Pre-vote negotiations in binary voting with non-manipulable rules

Grandi, Umberto; Grossi, Davide; Turrini, Paolo

Published in:
Journal of artificial intelligence research

DOI:
10.1613/jair.1.11446

Publication date:
2019

Journal of Artificial Intelligence Research 64 (2019) 895–929 Submitted 06/18; published 04/19

Negotiable Votes

Pre-Vote Negotiations in Binary Voting with Non-Manipulable Rules

Umberto Grandi UMBERTO.GRAND@IRIT.FR
Institut de Recherche en Informatique de Toulouse (IRIT)
University of Toulouse
2 rue du Doyen Gabriel-Marty, 31042 Toulouse, France

Davide Grossi D.GROSSI@RUG.NL
Bernoulli Institute for Mathematics,
Computer Science and Artificial Intelligence
University of Groningen
Bernoulliborg, Nijenborgh 9, 9747 AG Groningen, The Netherlands

Paolo Turrini P.TURRINI@WARWICK.AC.UK
Department of Computer Science
University of Warwick
CV4 7AL Coventry, United Kingdom

Abstract

We study voting games on binary issues, where voters hold an objective over the outcome of the collective decision and are allowed, before the vote takes place, to negotiate their ballots with the other participants. We analyse the voters' rational behaviour in the resulting two-phase game when ballots are aggregated via non-manipulable rules and, more specifically, quota rules. We show under what conditions undesirable equilibria can be removed and desirable ones sustained as a consequence of the pre-vote phase.

1. Introduction

Group decision-making is a topic of increasing relevance for Artificial Intelligence (AI). Addressing the problem of how groups of self-interested, autonomous entities can take the "right" decisions together is key to achieve intelligent behaviour in systems dependent on the interaction of autonomous entities. Against this backdrop, social choice theory has by now an established place in the toolbox of AI and, especially, multi-agent systems (henceforth, MAS), see, e.g., Shoham and Leyton-Brown (2008), Brandt, Conitzer, Endriss, Lang, and Procaccia (2016). Voting in particular has been extensively studied as a prominent group decision-making paradigm in MAS. Despite this, only a recent body of literature is starting to focus on voting as a form of strategic, non-cooperative, interaction, see, e.g., Desmedt and Elkind (2010), Xia and Conitzer (2010), Obraztsova, Markakis, and Thompson (2013), Meir, Lev, and Rosenschein (2014), Elkind, Grandi, Rossi, and Slinko (2015), Obraztsova, Rabinovich, Elkind, Polukarov, and Jennings (2016).[1]

1. This research direction was given particular momentum by the organisation of the Dagstuhl Workshop on Computation and Incentives in Social Choice in 2012, and the COST IC1205 Workshop on Iterative Voting and Voting Games, University of Padova, 2014.

More specifically—and this is the focus of the current contribution—no work with the notable exception of the literature on iterative voting (Meir, 2017, is a recent survey on this topic) has ever studied how voting behavior in rational agents is influenced by strategic forms of interaction that precede voting, like persuasion, or negotiation. Literature in social choice has recognised that interaction preceding voting can be an effective tool to induce opinion change and achieve compromise solutions (Dryzek & List, 2003; List, 2011) while in game theory pre-play negotiations are known to be effective in overcoming inefficient outcomes caused by players' individual rationality (Jackson & Wilkie, 2005).

1.1 Rational, Truthful, but Inefficient Votes

Consider a multiple referendum, where a group of voters cast yes/no opinions on a number of issues, which are then aggregated independently in order to obtain the group's opinion on those issues. An instance of this situation is represented in the following table:

	Issue 1	Issue 2	Issue 3
Voter 1	1	0	0
Voter 2	0	1	0
Voter 3	0	0	1
Majority	0	0	0

In the example above, three voters express their binary opinions on each of three issues (1 for acceptance, 0 for rejection), which are aggregated one by one using the majority rule. Voters typically approach such a referendum with some preference over its outcomes. So, enriching the example, let us assume that each voter i is interested in having the group accept issue i, and is indifferent about the remaining two issues. Given these goals, it is rational for each voter to cast a truthful vote, that is a vote in which each voter i accepts issue i: if the voter is not pivotal its vote will not count, but if it is pivotal, casting a truthful vote will make its own opinion become majority.

The example—which is also an instance of the so-called multiple election paradox (Brams, Kilgour, & Zwicker, 1998)—shows a situation in which truthful voting leads to an inefficient majority. That is, the outcome of the majority rule rejects all issues, and in so doing fails to meet the goal of each voter. Importantly, in the example above there are a number of profiles that would lead to an efficient majority (e.g., the profile where each voter accepts every issue). Even when sincere voting is rational, its outcome may turn out to be inefficient. As we will see, this is not a feature of the majority rule alone, but of a large class of well-behaved aggregation rules. Understanding mechanisms which can resolve such inefficiencies is, we argue, an important step towards the development of human-level group decision-making capabilities in artificial intelligence.

1.2 Contribution and Scientific Context

In this paper we study pre-vote negotiations in voting games over binary issues, where voters hold a simple type of lexicographic preference over the set of issues: they hold an objective about a subset of them while they are willing to strike deals on the remaining ones. Voters can influence one another before casting their ballots by transferring utility

in order to obtain a more favourable outcome in the end. We show that this type of pre-vote interaction has beneficial effects on voting games by refining their set of equilibria, and in particular by guaranteeing the efficiency of truthful ones. Specifically, we isolate precise conditions under which 'bad' equilibria—i.e., truthful but inefficient ones—can be overcome, and 'good' ones can be sustained. Our work relates directly to several on-going lines of research in social choice, game theory and their applications to MAS.

Binary Aggregation Aggregation and merging of information is a long studied topic in AI (Konieczny & Pino Peréz, 2002; Chopra, Ghose, & Meyer, 2006; Everaere, Konjeczny, & Marquis, 2007) and judgment aggregation has become an influential formalism in AI (Endriss, 2016). The basic setting of binary voting is also known as *voting in multiple referenda* (Lacy & Niou, 2000), and can be further enriched by imposing that individual opinions also need to satisfy a set of integrity constraints, like in binary voting with constraints (Grandi & Endriss, 2013) and judgment aggregation proper (Dietrich & List, 2007a; Grossi & Pigozzi, 2014). Standard preference aggregation, which is the classical framework for voting theory, is a special case of binary voting with constraints (Dietrich & List, 2007a). The introduction of constraints will be touched upon towards the end of the paper. Research in binary voting and judgment aggregation focused on the (non-)manipulability of judgment aggregation rules (Dietrich & List, 2007c; Botan, Novaro, & Endriss, 2016) and its computational complexity (Endriss, Grandi, & Porello, 2012; Baumeister, Erdélyi, Erdélyi, & Rothe, 2015), but a fully-fledged theory of non-cooperative games in this setting has not yet been developed and that is our focus here.

Election Control and Bribery The field of computational social choice has extensively studied decision problems that capture various forms of election control (see Faliszewski and Rothe 2016, for a recent overview) such as adding and deleting candidates, lobbying and bribery, modelled from the single agent perspective of a lobbyist or briber who tries to influence voters' decisions through monetary incentives, or from the perspective of a coalition of colluders (Bachrach, Elkind, & Faliszewski, 2011). Here we study a form of control akin to bribery, but where any voter can 'bribe' any other voter. Our work can be seen as an effort to develop a game-theoretic model of this type of control, and given our focus on equilibrium analysis we sidestep issues of computational complexity in this paper.

Equilibrium refinement Non-cooperative models of voting are known to suffer from a multiplicity of equilibria, many of which appear counterintuitive, not least because of their inefficiency. Equilibrium selection or refinement is a vast and long-standing research program in game theory (Meyerson, 1978). Models of equilibrium refinement have been applied to voting games in the literature on economics (Gueth & Selten, 1991; Kim, 1996) and within MAS especially within the above-mentioned iterative voting literature (Meir, 2017), which offers a natural strategy for selecting equilibria through the process of best response dynamics that starts from a profile of truthful votes. Our model tackles the same issue of refinement of equilibria in the context of binary voting, and focusing on those equilibria that are truthful and efficient. Unlike in iterative voting, our model is a two-phase model where equilibria are selected by means of an initial pre-vote negotiation phase, followed by voting.

Boolean Games We model voting strategies in binary aggregation with a model that generalises the well-known boolean games model (Harrenstein, van der Hoek, Meyer, & Witteveen, 2001; Wooldridge, Endriss, Kraus, & Lang, 2013): voters have control of a set of propositional variables, i.e., their ballot, and have specific goal outcomes they want to achieve. In our setting the goals of individuals are expressed *over the outcome* of the decision process, thus on variables that—in non-degenerate forms of voting—do not depend on their single choice only. Unlike boolean games, where each actor uniquely controls a propositional variable, in our setting the control of a variable is shared among the voters and its final truth value is determined by a voting rule. A formal relation with boolean games will be provided towards the end of the paper.

Pre-play Negotiations We model pre-vote negotiations as a pre-play interaction phase, in the spirit of Jackson and Wilkie (2005). During this phase, which precedes the play of a normal form game—the voting game— players are entitled to sacrifice part of their final utility in order to convince their opponents to play certain strategies, which in our case consist of voting ballots. In doing so we build upon the framework of endogenous boolean games (Turrini, 2016), which enriches boolean games with a pre-play phase. We abstract away from the sequential structure of the bargaining phase, modelled in closely related frameworks (Goranko & Turrini, 2016).

The use of side-payments for equilibrium refinement is not novel in Artificial Intelligence. Notably Monderer and Tennenholtz's (2004) k-implementation and "strong mediated equilibrium" (Monderer & Tennenholtz, 2009) employ side-payments by an external third party to drive the choices of self-interested agents. While both their models allow for game transformations through side-payments, the equilibrium refinement is mediated by an interested third party. Our model of side-payments and equilibrium refinement stems from Jackson and Wilkie's endogenous approach, which analyses the effect of negotiation through side-payments *without exogenous interventions*.

1.3 Outline of the Paper

The paper is organised as follows. In Section 2 we present the setting of binary aggregation, defining the (issue-wise) majority rule as well as a general class of aggregation procedures, which constitute the rules of choice for the current paper. In Section 3 we define voting games for binary aggregation, specifying individual preferences by means of both a goal and a utility function, and we show how undesirable equilibria can be removed by appropriate modifications of the game matrix. In Section 4 we present a full-blown model of collective decisions as a two-phase game, with a negotiation phase preceding voting. We show how the set of equilibria can be refined by means of rational negotiations. Section 5 relaxes assumptions we make on voters' goals in the basic framework showing the robustness of our results. Section 6 discusses related work in some more detail. Finally, Section 7 concludes.

2. Preliminaries: Binary Aggregation

The study of binary aggregation dates back to Wilson (1975), and was recently revived both within economics (Dokow & Holzman, 2010) and AI (Grandi & Endriss, 2013). This section is a brief introduction to the key notions and definitions of the framework.

	p	q	r
A	1	0	1
B	1	1	0
C	0	0	0
maj	1	0	0

Table 1: An instance of binary aggregation

2.1 Basic Definitions

In binary aggregation a finite set of agents (to which we will also refer as voters and, later on, players) $\mathcal{N} = \{1, \ldots, n\}$ express yes/no opinions on a finite set of binary issues $\mathcal{I} = \{1, \ldots, m\}$, and these opinions are then aggregated into a collective decision over each issue. This is analogous to voting over a simple (binary) combinatorial domain (Lang & Xia, 2016). Issues can be seen as queries to voters in a multiple referendum (Lacy & Niou, 2000), or seats that need to be allocated to candidates—two per seat—in assembly composition (Benôit & Kornhauser, 2010), or candidates of which voters approve or disapprove of (Brams & Fishburn, 1978).

In this paper we assume that $|\mathcal{N}| \geq 3$, i.e., there are at least 3 individuals. We denote by $\mathcal{D} = \{B \mid B : \mathcal{I} \to \{0, 1\}\}$ the set of all possible binary opinions over the set of issues $\mathcal{I}$ and call an element $B \in \mathcal{D}$ a **ballot**. $B(j) = 0$ (respectively, $B(j) = 1$) indicates that the agent who submits ballot B rejects (respectively, accepts) the issue j. A **profile** $\boldsymbol{B} = (B_1, \ldots, B_n) \in \mathcal{D}^{\mathcal{N}}$ is the choice of a ballot for every individual in $\mathcal{N}$. Given a profile $\boldsymbol{B}$, we use B_i to denote the ballot of individual i within a profile $\boldsymbol{B}$. We adopt the usual convention writing $-i$ for $\mathcal{N} \setminus \{i\}$ and thus $\boldsymbol{B}_{-i}$ to denote the sequence consisting of the ballots of individuals other than i. Thus, $B_i(j) = 1$ indicates that individual i accepts issue j in profile $\boldsymbol{B}$. Furthermore, we denote by $\mathcal{N}_j^{\boldsymbol{B}} = \{i \in \mathcal{N} \mid B_i(j) = 1\}$ the set of individuals accepting issue j in profile $\boldsymbol{B}$.

Given a set of individuals $\mathcal{N}$ and issues $\mathcal{I}$, an aggregation rule or **aggregator** for $\mathcal{N}$ and $\mathcal{I}$ is a function $F : \mathcal{D}^{\mathcal{N}} \to \mathcal{D}$, mapping every profile to a binary ballot in $\mathcal{D}$, called the **collective** ballot. $F(\boldsymbol{B})(j) \in \{0, 1\}$ denotes the value of issue j in the aggregation of $\boldsymbol{B}$ via aggregator F. The benchmark aggregator is the so-called **issue-wise strict majority rule**, which we will refer to simply as (issue-wise) majority (in symbols, maj). The rule accepts an issue if and only if a majority of voters accept it, formally $\mathsf{maj}(\boldsymbol{B})(j) = 1$ if and only if $|\mathcal{N}_j^{\boldsymbol{B}}| \geq \frac{|\mathcal{N}|+1}{2}$. Table 1 depicts a ballot profile with three agents (rows A, B and C) on three issues (columns p, q and r) and the aggregated ballot (fourth row) obtained by majority.

Other examples of aggregation rules include quota rules, which we discuss below, and degenerate rules such as dictatorships, for which there exists $i \in \mathcal{N}$ such that on every issue j, and for all profile $\boldsymbol{B}$, $F(\boldsymbol{B})(j) = B_i(j)$.

2.2 Types of Aggregators of Interest

In this paper we focus specifically on two classes of aggregators: the class of non-manipulable aggregators, an its subclass consisting of all quota rules.

2.2.1 Non-Manipulable Aggregators

In binary aggregation, an aggregator F is said to be **non-manipulable** if there exists no profile $\boldsymbol{B}$ such that for some issue $j \in \mathcal{I}$ and agent $i \in \mathcal{N}$, $B_i(j) \neq F(\boldsymbol{B})(j)$ and $B_i(j) = F(B'_i, \boldsymbol{B}_{-i})(j)$ for some ballot $B'_i \neq B_i$ (Dietrich & List, 2007c). That is, no agent accepting (resp., rejecting) an issue while the issue is collectively rejected (resp., accepted) can change its ballot in order to force the issue to be collectively accepted (resp., rejected).[2]

The class of non-manipulable aggregators is the baseline class for the analysis of binary voting games developed in the paper. This is, we argue, a natural class of aggregators to focus on for our purposes, as we will be concerned with the analysis of strategic behavior that arises in a pre-vote negotiation phase. In a sense our models show how rich strategic behavior can be supported even by aggregators that make vote manipulation impossible.

It is known (Dietrich & List, 2007c, Th. 1) that the class of non-manipulable aggregators corresponds to the class of aggregators which are independent and monotonic, so we will be referring to the two classes interchangeably. An aggregator F is said to be: **independent** if for all issue $j \in \mathcal{I}$ and any two profiles $\boldsymbol{B}, \boldsymbol{B'} \in \mathcal{D}^{\mathcal{N}}$, if $B_i(j) = B'_i(j)$ for all $i \in \mathcal{N}$, then $F(\boldsymbol{B})(j) = F(\boldsymbol{B'})(j)$; **monotonic** if for all issue $j \in \mathcal{I}$, $x \in \{0, 1\}$ and any two profiles $\boldsymbol{B}, \boldsymbol{B'} \in \mathcal{D}^{\mathcal{N}}$, if $B_i(j) = x$ entails $B'_i(j) = x$ for all $i \in \mathcal{N}$, and for some $\ell \in \mathcal{N}$ we have that $B_\ell(j) = 1 - x$ and $B'_\ell(j) = x$, then $F(\boldsymbol{B})(j) = x$ entails $F(\boldsymbol{B'})(j) = x$. Intuitively, an aggregator is independent if the decision of accepting a given issue j does not depend on the judgment of the individuals on any issue other than j.[3] It is monotonic if increasing (respectively, decreasing) the support on one issue when this is collectively accepted (respectively, rejected), does not modify the result.

The class of non-manipulable aggregators includes the majority rule, as well as any quota rule, to which we will turn next. The minority rule, i.e., the rule that always selects the opposite of the majority rule, satisfies independence, but fails monotonicity. Dictatorships, oligarchies, and similar non-anonymous aggregation rules[4] do satisfy independence and monotonicity. Rules that fall out of the class, and that have been studied in the literature, are also the so-called distance-based rules, which output the ballot that minimises the overall distance to the profile for a suitable notion of distance.[5]

2.2.2 Quota Rules

While the core of our results are proven for independent and monotonic aggregators, we will sometimes restrict ourselves to the class of quota rules and establish stronger claims for that class. Quota rules accept an issue if the number of voters accepting it exceeds a given positive quota, possibly different for each issue. Formally, if a quota rule F_q is defined via a function $q : \mathcal{I} \to \{1, \ldots, n\}$, associating a quota to each issue, by stipulating

2. This notion of non-manipulability is a preference-free variant of the one used in the preference aggregation literature.
3. When the aggregator is independent, the process of aggregation is also referred to as proposition-wise voting in the literature on multiple referrenda (Ozkai-Sanver & Sanver, 2006), or seat-by-seat voting in the literature on assembly composition (Benôit & Kornhauser, 2010), or as simultaneous voting in the literature on voting over combinatorial domains (Lang & Xia, 2016).
4. See definition of anonymity below.
5. See Grossi and Pigozzi (2014) and Endriss (2016) for a more detailed exposition of aggregation rules.

$F_q(\boldsymbol{B})_j = 1 \Leftrightarrow |N_j^{\boldsymbol{B}}| \geq q(j)$.[6] F_q is called **uniform** in case q is a constant function. Issue-wise majority is a uniform quota rule with $q = \left\lceil \frac{|\mathcal{N}|+1}{2} \right\rceil$.

In our results we will make use of known axiomatic characterizations of quota rules. The class of quota rules corresponds to the class of independent, monotonic, responsive and anonymous aggregators (Dietrich & List, 2007b, Theorem 1).[7] The class of uniform quota rules corresponds to the class of aggregators that are, in addition, neutral (Endriss, 2016, Corollary 17.5). An aggregator F is said to be: **neutral** if for any two issues $j, k \in \mathcal{I}$ and any profile $\boldsymbol{B} \in \mathcal{D}$, if for all $i \in \mathcal{N}$ we have that $B_i(j) = B_i(k)$, then $F(\boldsymbol{B})(j) = F(\boldsymbol{B})(k)$;[8] **responsive** if for all issue $j \in \mathcal{I}$ there exist two profiles $\boldsymbol{B}$ and $\boldsymbol{B'}$ such that $F(\boldsymbol{B})(j) = 1$ and $F(\boldsymbol{B'})(j) = 0$; **anonymous** if for all two players $i, j \in \mathcal{N}$ and any two profiles $\boldsymbol{B}, \boldsymbol{B'} \in \mathcal{D}^{\mathcal{N}}$, if $B_i = B'_j$, $B'_i = B_j$ and, for each $k \in \mathcal{N} \backslash \{i, j\}$, we have that $B_k = B'_k$, then $F(\boldsymbol{B}) = F(\boldsymbol{B'})$. Intuitively, an aggregator is responsive if it is not a constant function. It is neutral if all issues are treated in the same way, and anonymous if all voters are treated in the same way.

2.3 Winning and Veto Coalitions

Given an aggregator F, we call a set of voters $C \subseteq \mathcal{N}$ a **winning coalition** for issue $j \in \mathcal{I}$ if for every profile $\boldsymbol{B}$ we have that *if* $B_i(j) = 1$ for all $i \in C$ and $B_i(j) = 0$ for all $i \notin C$ *then* $F(\boldsymbol{B})(j) = 1$. [9] The notion of winning coalition is closely related to the independence property defined above:[10]

Fact 1. *An aggregator F is independent if and only if for all $j \in \mathcal{I}$ there exists a set of subsets $\mathcal{W}_j \subseteq \mathcal{P}(\mathcal{N})$ such that, for each ballot $\boldsymbol{B}$, $F(\boldsymbol{B})_j = 1$ if and only $N_j^{\boldsymbol{B}} \in \mathcal{W}_j$.*

That is, F is independent if and only if it can be defined in terms of a family $\{\mathcal{W}_j\}_{j \in \mathcal{I}}$ of sets of winning coalitions. Furthermore, we call a set of voters $C \subseteq \mathcal{N}$ a **veto coalition** for issue $j \in \mathcal{I}$ if for every profile $\boldsymbol{B}$ we have that *if* $B_i(j) = 0$ for all $i \in C$ and $B_i(j) = 1$ for all $i \notin C$ *then* $F(\boldsymbol{B})(j) = 0$. Clearly,

$$C \in \mathcal{W}_j \text{ IFF } \mathcal{N} \backslash C \notin \mathcal{V}_j. \tag{1}$$

Observe that by the above relation and Fact 1 an independent aggregator can equivalently be defined by a set $\{\mathcal{V}_j\}_{j \in \mathcal{I}}$ of veto coalitions, consisting of the complements of those coalitions that do not belong to $\mathcal{W}_j$. Let us fix intuitions by a few examples. The sets of winning and veto coalitions for issue-wise majority are, for every issue j: $\mathcal{W}_j = \left\{ C \subseteq \mathcal{N} \;\middle|\; |C| \geq \frac{|\mathcal{N}|+1}{2} \right\}$ and $\mathcal{V}_j = \left\{ C \subseteq \mathcal{N} \;\middle|\; |C| \geq \frac{|\mathcal{N}|}{2} \right\}$. When $|\mathcal{N}|$ is odd, this is the only quota rule for which $\mathcal{W}_j = \mathcal{V}_j$.[11] For a constant aggregator which always accepts all issues, that is, a quota rule

6. Note that we exclude from our definition of quota rules constant functions, that is, quota rules with quota equal to 0 or larger than n.
7. See also Endriss (2016), Proposition 17.4.
8. Independent and neutral aggregators are often called systematic.
9. The definitions from this section are well-known from the literature, except for the notion of veto and resilient coalitions treated below. See, for instance, Dokow and Holzman (2010).
10. This and the following facts are assumed to be well-known, for a proof in the setting of judgment aggregation we refer to Endriss (2016), Lemma 17.1.
11. This property is known as unbiasedness.

with $q = 0$, the sets of winning and veto coalitions are, for each issue j: $\mathcal{W}_j = 2^{\mathcal{N}}$, i.e., any coalition is winning, and $\mathcal{V}_j = \emptyset$, i.e., no coalition is a veto coalition.

Additional properties imposed on an independent aggregator F induce further structure on winning (and veto) coalitions:

Fact 2. *An independent aggregator F is monotonic if and only if for each $j \in \mathcal{I}$ and for any $C \in \mathcal{W}_j$ (respectively, $C \in \mathcal{V}_j$), if $C \subseteq C'$ then $C' \in \mathcal{W}_j$ (resp., $C' \in \mathcal{V}_j$), i.e., winning (and veto) coalitions are closed under supersets. It is neutral if and only if for each $j, k \in \mathcal{I}$ we have that $\mathcal{W}_j = \mathcal{W}_k$ (equivalently, $\mathcal{V}_j = \mathcal{V}_k$ for all $j, k \in \mathcal{I}$). It is anonymous if and only if for each $j \in \mathcal{I}$ we have that $C \in \mathcal{W}_j$ (resp., $C \in \mathcal{V}_j$) implies that $D \in \mathcal{W}_j$ (resp., $D \in \mathcal{V}_j$) whenever $|C| = |D|$, i.e., coalitions are winning (resp., veto) only depending on their cardinality.*

We conclude this preliminary section with one last important definition. We call C a **resilient winning coalition** (respectively, a **resilient veto coalition**) for issue $j \in \mathcal{I}$ if C is a winning (resp., veto) coalition for j and, for every $i \in C$, $C \setminus \{i\}$ is also a winning (resp., veto) coalition for j.[12] For every issue j, the set of resilient winning (resp., veto) coalitions for j is denoted $\mathcal{W}_j^+$ (resp., $\mathcal{V}_j^+$).

Given a neutral aggregator F, such as a uniform quota rule, we denote with $\mathcal{W}$ (resp., $\mathcal{V}$) the set of winning (resp., veto) coalitions for F, and with $\mathcal{W}^+$ (resp., $\mathcal{V}^+$) the set of resilient winning (resp., veto) coalitions for F. In the case of the majority rule we have $\mathcal{W}^+ = \left\{C \subseteq \mathcal{N} \mid |C| \geq \frac{|\mathcal{N}|+1}{2} + 1\right\}$, i.e., all winning coalitions exceeding the majority threshold of at least one element are resilient winning coalitions. For a uniform quota rule F_q, the set of resilient winning coalitions is $\mathcal{W}^+ = \{C \subseteq \mathcal{N} \mid |C| \geq q + 1\}$. Observe also that if F is dictatorial on every issue, then $\mathcal{W}^+ = \emptyset$.

3. Games for Binary Aggregation

In this section we present a model of a strategic interaction played among voters involved in a collective decision-making problem on binary issues. Players' strategies consist of all binary ballots on the set of issues, and the outcome of the game is obtained by aggregating the individual ballots by means of a given aggregator. Players' preferences are expressed in the form of a simple goal that is interpreted on the outcomes of the aggregation (i.e., the collective decision), and by an explicit payoff function for each player i, yielding to player i a real number at each profile and encoding, intuitively, the utility i would receive, should that profile of votes occur. We study the existence of equilibria of these games, paying particular attention to the truthful and efficient ones.

3.1 Main Definitions

Before defining aggregation games we need a last piece of notation. To each set of issues $\mathcal{I}$, we associate the set of propositional atoms $PS = \{p_1, \ldots, p_{|\mathcal{I}|}\}$ containing one atom for each issue in $\mathcal{I}$. We denote by $\mathcal{L}_{PS}$ the propositional language constructed by closing PS under a functionally complete set of boolean connectives (e.g., $\{\neg, \wedge\}$).

12. This definition adapts the notion of resiliency of equilibria studied by Halpern (2011) to the notion of winning and veto coalition proper of binary aggregation.

3.1.1 Aggregation Games, Goals and Preferences

Definition 1. *Let $\mathcal{I}$ and $\mathcal{N}$ be given. An* aggregation game *(for $\mathcal{I}$ and $\mathcal{N}$) is a tuple $\mathcal{A} = \langle \mathcal{N}, \mathcal{I}, F, \{\gamma_i\}_{i\in\mathcal{N}}, \pi \rangle$ where:*

- *F is an aggregator for $\mathcal{N}$ and $\mathcal{I}$;*
- *each γ_i is a cube, i.e. a conjunction of literals from $\mathcal{L}_{PS}$,*[13] *which is called a* goal*;*
- *$\pi : \mathcal{N} \to \left(\mathcal{D}^{\mathcal{N}} \to \mathbb{R}\right)$ is a payoff function assigning to each agent and each strategy profile a real number representing the utility that player i gets at that profile. For each player i, the payoff function $\pi(i)$ of player i will be denoted simply by π_i.*

Note that a strategy profile in an aggregation game is a profile of binary ballots, and will therefore be denoted with $\boldsymbol{B}$. In the context of aggregation games we will use the term "strategy profile" and "ballot profile", or even just "profile", interchangeably.

Goals, intuitively, represent properties of the outcome of the aggregation process that voters are not willing to compromise about. By making the assumptions that goals are cubes we assume that each voter has a simple incentive structure: she can identify a specific set of atoms that she wants to be true at the outcome, another set of atoms that she wants to be false, and she is indifferent about the value of the others. When comparing two outcomes, one of which satisfies her goal and one of which does not, a voter will choose the outcome satisfying her goal. Thus, the first degree of preference of agents is dichotomous (Elkind & Lackner, 2015). If then two outcomes both satisfy her goal, or both do not, then the voter will look at the value she obtains at those outcomes through her payoff function.

This, we argue, is a very natural class of preferences for binary aggregation. They are technically known as **quasi-dichotomous** preferences and have been studied in the context of Boolean games (Wooldridge et al., 2013). Henceforth we employ the satisfaction relation $\models$ (respectively, its negation $\not\models$) to express that a ballot satisfies (respectively, does not satisfy) a goal. The preference relation induced on ballot profiles by goals and payoff functions is defined as follows:

Definition 2 (Quasi-dichotomous preferences)**.** *Let $\mathcal{A}$ be an aggregation game. Ballot profile $\boldsymbol{B}$ is strictly preferred by $i \in \mathcal{N}$ over ballot profile $\boldsymbol{B}'$ (in symbols, $\boldsymbol{B} \succ_i^{\pi} \boldsymbol{B}'$) if and only if any of the two following conditions holds:*

i) $F(\boldsymbol{B}') \not\models \gamma_i$ and $F(\boldsymbol{B}) \models \gamma_i$;

ii) $F(\boldsymbol{B}') \models \gamma_i$ if and only if $F(\boldsymbol{B}) \models \gamma_i$, and $\pi_i(\boldsymbol{B}) > \pi_i(\boldsymbol{B}')$.

The two ballot profiles are equally preferred whenever $F(\boldsymbol{B}') \models \gamma_i$ if and only if $F(\boldsymbol{B}) \models \gamma_i$, and $\pi_i(\boldsymbol{B}) = \pi_i(\boldsymbol{B}')$. The resulting weak preference order among ballot profiles is denoted $\succeq_i^{\pi}$.

In other words, a profile $\boldsymbol{B}$ is weakly preferred by player i to $\boldsymbol{B}'$ if either $F(\boldsymbol{B})$ satisfies i's goal and $F(\boldsymbol{B}')$ does not or, if both satisfy i's goal or neither do, but $\boldsymbol{B}$ yields to i at

13. Formally, each γ_i is equivalent to $\bigwedge_{j\in K} \ell_j$ where $K \subseteq \mathcal{I}$ and $\ell_j = p_j$ or $\ell_j = \neg p_j$ for all $j \in K$.

least as good a payoff as $\boldsymbol{B}'$. Individual preferences over ballot profiles are therefore induced by their goals, by their payoff functions, and by the aggregation procedure used.

Finally, goals relate in a clear way to the structure of winning and veto coalitions of an independent aggregator. One can identify for every goal which are the sets of agents that can force the acceptance of such goal:

Definition 3. *For an independent aggregator F and a cube γ we say that C is* winning for *γ if and only if $C \in \mathcal{W}_j$ for each j such that γ logically entails p_j, and $C \in \mathcal{V}_j$ for each j such that γ logically entails $\neg p_j$. We write $\mathcal{W}_\gamma$ for the set of winning coalitions for γ. The set of resilient winning coalitions for γ (denoted $\mathcal{W}^+_\gamma$) is defined in the obvious way.*

In words, a coalition is winning for a goal whenever it is winning for all the issues that need to be accepted for γ to be satisfied, and veto for all the issues that need to be rejected for γ to be true. An obvious adaptation of the definition yields the notion of veto coalition for a given goal.

3.1.2 Classes of Aggregation Games

A natural class of aggregation games is that of games where the individual utility only depends on the outcome of the collective decision:

Definition 4. *An aggregation game $\mathcal{A}$ is called* **uniform** *if for all $i \in \mathcal{N}$ and profiles $\boldsymbol{B}$ it is the case that $\pi_i(\boldsymbol{B}) = \pi_i(\boldsymbol{B}')$ whenever $F(\boldsymbol{B}) = F(\boldsymbol{B}')$. A game is called* **constant** *if all π_i are constant functions, i.e., for all $i \in \mathcal{N}$ and all profiles $\boldsymbol{B}$ we have that $\pi_i(\boldsymbol{B}) = \pi_i(\boldsymbol{B}')$.*

Clearly, all constant aggregation games are uniform. Games with uniform payoffs are arguably the most natural examples of aggregation games. The payoff each player receives is only dependent on the outcome of the vote, and not on the ballot profile that determines it. For convenience, we assume that in uniform games the payoff function is defined directly on outcomes, i.e., $\pi_i : \mathcal{D} \to \mathbb{R}$. Constant games are games where players' preferences are fully defined by their goals, and are therefore dichotomous.

We call a strategy B *i*-**truthful** if it satisfies the individual's goal γ_i. Note that in the case in which γ_i is a cube that specifies in full a single binary ballot—that is, the agent's goal *is* a ballot—our notion of truthfulness coincides with the classic notion used in judgment aggregation and binary voting where only one ballot is truthful, and all other ballots are available for strategic voting.

Let us introduce some further terminology concerning strategy profiles:

Definition 5. *Let $C \subseteq \mathcal{N}$. We call a strategy profile $\boldsymbol{B} = (B_1, \ldots, B_n)$:*

(1) C-truthful if all B_i with $i \in C$ are i-truthful, i.e., $B_i \models \gamma_i$, for all $i \in C$;

(2) C-goal-efficient (C-efficient) if $F(\boldsymbol{B}) \models \bigwedge_{i \in C} \gamma_i$;[14]

(3) totally C-goal-inefficient (totally C-inefficient) if $F(\boldsymbol{B}) \models \bigwedge_{i \in C} \neg\gamma_i$.

*An aggregation game is called C-***consistent***, for $C \subseteq \mathcal{N}$, if the conjunction of the goals of agents in coalition C is consistent, i.e., if the formula $\bigwedge_{i \in C} \gamma_i$ is satisfiable.*

14. Observe that since each γ_i is a cube, $\bigwedge_{i \in C} \gamma_i$ is also a cube.

Observe that while the notion of truthfulness is a property of the ballot itself, with goals interpreted on the individual ballots to check for truthfulness, the two notions of efficiency are instead properties of the outcome of the aggregation.

3.2 Equilibria in Aggregation Games

In this section we study Nash equilibria (NE) in aggregation games and specific classes thereof. Our focus is the existence of 'good' NE, that is, NE that are truthful and efficient.

3.2.1 Inexistence of Equilibria in (General) Aggregation Games

First of all, it is important to observe that aggregation games may not have, in general, (pure stragegy) NE.

Fact 3. *There are aggregation games that have no NE.*

Proof. Define an aggregation game as follows. Let $\mathcal{I} = \{p\}$, $\mathcal{N} = \{1, 2, 3\}$, and let $\gamma_1 = \gamma_2 = \gamma_3 = \top$. Assume also that the payoff function is defined as follows: for any ballot profile $\boldsymbol{B}$, we have that $\pi_1(\boldsymbol{B}) = 1$ if and only if $B_1 = B_2$, and 0 otherwise; $\pi_2(\boldsymbol{B}) = 1$ if and only if $B_1 \neq B_2$, and 0 otherwise; finally, π_3 is constant. That is, agent 1 wants 1 and 2 to agree on issue p while agent 2 wants them to disagree, and agent 3 is indifferent among any two outcomes of the interaction.[15] It is easy to see that the aggregation game encodes a matching-pennies type of game between 1 and 2 and, therefore, the resulting aggregation game does not have a NE. □

We will come back to the issue of the inexistence of NE in Section 4.

3.2.2 Equilibria in Constant Aggregation Games

Recall that a strategy B_i is **weakly dominant** for agent i if for all profiles $\boldsymbol{B}$ we have that $(\boldsymbol{B}_{-i}, B_i) \succeq_i^{\pi} \boldsymbol{B}$. We begin with an important result showing that in constant aggregation games with aggregators that are non-manipulable (i.e., independent and monotonic), the truthfulness of a strategy is a sufficient condition for it to be weakly dominant.

Proposition 4. *Let $\mathcal{A}$ be a constant aggregation game with F non-manipulable, and let $i \in \mathcal{N}$ be a player. If a strategy B_i is truthful then it is weakly dominant for i.*[16]

Proof. Let B_i be a truthful strategy, i.e., $B_i \models \gamma_i$. We want to show that B_i is weakly dominant, that is for every $\boldsymbol{B}' \in \mathcal{D}^{\mathcal{N}}$, $F(\boldsymbol{B}') \models \gamma_i$ implies $F(\boldsymbol{B}'_{-i}, B_i) \models \gamma_i$. We proceed towards a contradiction and assume that, for some profile $\boldsymbol{B}'$ we have that $F(\boldsymbol{B}') \models \gamma_i$ and $F(\boldsymbol{B}'_{-i}, B_i) \not\models \gamma_i$. Since by Definition 1 individual goals are cubes, we have that $\gamma_i = \bigwedge_{j \in \mathcal{I}} \ell_j$, where ℓ_j is a literal built from *PS*. Hence there exists a $k \in \mathcal{I}$ such that $F(\boldsymbol{B}'_{-i}, B_i) \not\models \ell_k$ but $F(\boldsymbol{B}') \models \ell_k$. Assume w.l.o.g. that ℓ_k is positive, i.e., $\ell_k = p_k$. Since B_i is assumed to be truthful, $B_i \models \ell_k$ (that is, $B_i(k) = 1$). Now, F is independent so the value of issue k in the output of F depends only on the values of k in each individual ballot in the input profile. Moreover, since $F(\boldsymbol{B}') \models \ell_k$ and $B_i \models \ell_k$, by the monotonicity of F we conclude that $F(\boldsymbol{B}'_{-i}, B_i) \models \ell_k$. Contradiction. □

15. Note that this game is not uniform.

16. This proposition can also be obtained as corollary of a result by Dietrich and List (2007c). We are indebted to Ulle Endriss for this observation.

Intuitively the proposition tells us that independent and monotonic aggregators, as far as only the satisfaction of individual goals is concerned, guarantee that players are always better off by casting a truthful ballot. A first immediate consequence is that computing weakly dominant strategies in constant aggregation games takes a polynomial amount of time, since it boils down to finding a satisfying assignment to the individual goal, which in our model is a conjunction of literals. Other consequences are stated in the following corollary:

Corollary 5. *Let $\mathcal{A}$ be an aggregation game with F non-manipulable:*

(i) any profile $\boldsymbol{B}$ such that $B_i \models \gamma_i$ for all $i \in \mathcal{N}$ is a NE;

(ii) if for all $i \in \mathcal{N}$ the formula γ_i is consistent, then $\mathcal{A}$ has at least one NE.

For the subclass of non-manipulable aggregators consisting of quota rules, the converse of Proposition 4 holds, showing that for quota rules the truthfulness of a strategy is also a necessary condition for it to be weakly dominant.

Proposition 6. *Let $\mathcal{A}$ be a constant aggregation game with F a quota rule. Then all weakly dominant strategies for $i \in \mathcal{N}$ are truthful.*

Proof. Recall that quota rules correspond to the class of independent, monotonic, responsive and anonymous aggregators. Consider a weakly dominant strategy B_i for i, and assume towards a contradiction that $B_i \not\models \gamma_i$, i.e., B_i is not truthful. Since goals are cubes, there exists a k such that $B_i \not\models \ell_k$, where $\gamma_i = \bigwedge_{j \in \mathcal{I}} \ell_j$. W.l.o.g., assume that $\ell_k = p_k$ and let us argue towards the acceptance of k. By responsiveness there exist $\boldsymbol{B}$ and $\boldsymbol{B'}$ such that $F(\boldsymbol{B})(k) = 1$ and $F(\boldsymbol{B'})(k) = 0$. By anonymity, collective acceptance and rejection of k depends only on the cardinality of the set of agents accepting k in a given profile. So, by monotonicity there exists an integer q ($\neq 0$, by responsiveness) such that F accepts k if and only if the cardinality of the set of agents accepting k is at least q. Therefore there exist $\boldsymbol{B}$ and $\boldsymbol{B'}$ such that $F(\boldsymbol{B})(k) = 1$ and $F(\boldsymbol{B'})(k) = 0$ and such that $\boldsymbol{B}_{-i} = \boldsymbol{B'}_{-i}$ and $B_i(k) = 1 \neq B'_i(k)$. Furthermore, exploiting independence, let us assume that for any $j \neq i$ and $t \neq k$, $B_j(t) = B'_j(t)$ and $F(\boldsymbol{B}) \models \gamma_i$. That is, there exist two profiles such that the first satisfies i's goal and the second does not, because the first meets the threshold for accepting k while the second does not. And the reason for the second not meeting the threshold is i's ballot to reject k. In other words, i is pivotal in $\boldsymbol{B'}$ for the acceptance of k and therefore for the acceptance of its goal. It follows that for any strategy B'_i such that $B'_i \models \ell_k$ we have that $F(\boldsymbol{B}_{-i}, B'_i) \models \gamma_i$, i.e., B_i is dominated by B'_i, against the assumption that B_i is weakly dominant in $\mathcal{A}$. □

So quota rules are a subclass of non-manipulable aggregators for which weak dominance and truthfulness are equivalent conditions on players' strategies:

Corollary 7. *Let $\mathcal{A}$ be a constant aggregation game with F a quota rule. A strategy B_i is weakly dominant if and only if it is i truthful.*

It is worth to observe that Proposition 4 ceases to hold if we allow the goals of the voters to be propositional formulas more complex than a cube:

Example 1. *Let $F = \mathsf{maj}$, $\mathcal{N} = \{1,2,3\}$ and let $\mathcal{I} = \{1,2\}$. Let then $\gamma_1 = p_1 \vee p_2$ and $\gamma_2 = \gamma_3 = \top$. That is, agent 1 is interested in having at least one of the two issues accepted, while the rest of the agents are indifferent. We show that in this game not all truthful ballots of 1 are weakly dominant. Consider the profile $\boldsymbol{B} = (B_1, B_2, B_3)$ where 1 votes the truthful ballot $B_1 = (0,1)$, 2 votes $B_2 = (0,0)$ and 3 votes $B_3 = (1,0)$. We have that $F(\boldsymbol{B}) \not\models \gamma_1$. Clearly, 1 has a best response $B_1' = (1,0) \neq B_1$ in that profile as $F(B_1', B_2, B_3) \models \gamma_1$.*

Example 2. *Let $F = \mathsf{maj}$, $\mathcal{N} = \mathcal{I} = \{1,2,3\}$ and let agent 1's goal be that of having an odd number of accepted issues, while agents 2 and 3 have no specific goals. Formally, let $\gamma_1 = (p_1 \wedge p_2 \wedge p_3) \vee (p_1 \wedge \neg p_2 \wedge \neg p_3) \vee (\neg p_1 \wedge p_2 \wedge \neg p_3) \vee (\neg p_1 \wedge \neg p_2 \wedge p_3)$. As above we show that not all truthful ballots of 1 are weakly dominant. Consider the profile $\boldsymbol{B} = (B_1, B_2, B_3)$ where $B_2 = (1,0,0)$ and $B_3 = (0,1,0)$ and where 1 votes a truthful ballot $B_1 = (0,0,1)$. This profile results under the majority rule in $(0,0,0)$. So the (non-truthful) ballot $B_1' = (1,0,1) \neq B_1$ is a better response in $\boldsymbol{B}$ for 1 and yields the collective ballot $(1,0,0)$, which satisfies γ_1.*[17]

In fact, cubes guarantee that players' preferences in constant aggregation games satisfy a property known as separability,[18] which has been shown in other voting frameworks to guarantee that truthful strategies are undominated (Benôit & Kornhauser, 2010).[19]

3.2.3 Equilibria, Truthfulness and Efficiency

Proposition 4 establishes the existence of truthful equilibria. However, as the example in the introduction has shown, truthfulness does not guarantee efficiency in constant aggregation games. We show now that this does not only hold for constant aggregation games based on majority, but for constant aggregation games based on uniform quota rules in general:

Proposition 8. *For every uniform quota rule, there exist constant aggregation games with truthful and totally inefficient NE in weakly dominant strategies.*

Proof. Recall that uniform quota rules correspond to the class of aggregators which are independent, monotonic, neutral, anonymous and responsive.[20] The proof is by construction of a constant aggregation game $\mathcal{A}$ with the desired property. Let F be anonymous, systematic and monotonic, let $\mathcal{N} = \mathcal{I} = \{1,2,3\}$. First we will construct two games and show that at least one of the two admits a truthful and totally inefficient profile. We can then apply Proposition 4 to show that such profile ought to be a NE in weakly dominant strategies. The games are built on the same aggregator F and differ only on their goals. Game A Let, for each $i \in \mathcal{N}$, $\gamma_i = p_i$. Each γ_i is therefore, trivially, a cube. Note also

17. We are indebted to Edith Elkind for this example.
18. Cf. Lang and Xia (2016) for separability in combinatorial domains.
19. Separability is normally defined over strict orders (but cf. Hodge (2002) for a general treatment of the notion). In our setup separability can be defined as follows: $\preceq_i$ is separable if and only if for all $j \in \mathcal{I}$, if $\boldsymbol{B} \preceq_i \boldsymbol{B}'$ for two profiles such that $F(\boldsymbol{B})(k) = F(\boldsymbol{B}')(k)$ for all $k \neq j$, then $\boldsymbol{B}'' \preceq_i \boldsymbol{B}'''$ for all profiles such that $F(\boldsymbol{B}'')(k) = F(\boldsymbol{B}''')(k)$ for all $k \neq j$, $F(\boldsymbol{B})(j) = F(\boldsymbol{B}'')(j)$, and $F(\boldsymbol{B}')(j) = F(\boldsymbol{B}''')(j)$. If $\preceq_i$ is a dichotomous preference induced by a cube (i.e., a conjunction of literals), then $\preceq_i$ is separable under the above definition.
20. Responsiveness does not play a role in the proof and could be dispensed with. Concretely, this means that the claim holds also for trivial quota rules, that is, constant aggregators.

that $\bigwedge_{i\in\mathcal{N}}\gamma_i$ is satisfiable, so Game A is $\mathcal{N}$-consistent. Game B Let, for each $i \in \mathcal{N}$, the goal be defined as $\chi_i = \neg p_i$. Each χ_i is therefore, again, a trivial cube and Game B is also $\mathcal{N}$-consistent. Now construct a ballot profile $\boldsymbol{B}$ in Game A and a ballot profile $\boldsymbol{B'}$ in Game B as follows, for $i \in \mathcal{N}$ and $j \in \mathcal{I}$:

$$B_i(j) = \begin{cases} 1 & \text{if } p_j = \gamma_i \\ 0 & \text{otherwise} \end{cases} \qquad B'_i(j) = \begin{cases} 0 & \text{if } \neg p_j = \chi_i \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

That is, in $\boldsymbol{B}$ each voter votes 1 only on the issues which coincides with its goal. Vice versa, in $\boldsymbol{B'}$ each voter votes 0 only on the issues whose rejection coincides with its goal. By construction, $\boldsymbol{B}$ and $\boldsymbol{B'}$ are both truthful. Now assume that $\boldsymbol{B}$ is not totally inefficient, and we prove that $\boldsymbol{B'}$ is totally inefficient. So there exists $i \in \mathcal{N}$ such that $F(\boldsymbol{B}) \models \gamma_i$ and since by construction $\gamma_i = p_i$, $F(\boldsymbol{B})(i) = 1$. By anonymity and systematicity, $\boldsymbol{B}$ must actually be efficient, that is $F(\boldsymbol{B}) \models \bigwedge_{i\in\mathcal{N}}\gamma_i$ as by construction the individual positions on each issue are identical, modulo permutations. More precisely, by construction for all $j \neq i \in \mathcal{N}$, $B_j(p_i) = 0$, that is, i is the only voter accepting p_i in $\boldsymbol{B}$. By independence and monotonicity it follows that $\{i\} \in \mathcal{W}_{p_i}$, that is, if i accepts p_i so does F. However, $\{i\} \notin \mathcal{V}_{p_i}$ for otherwise i would be a dictator for p_i, against the assumption of anonymity for F. From $\{i\} \notin \mathcal{V}_{p_i}$ and Equation (1) it follows that $\mathcal{N}\setminus\{i\} \in \mathcal{W}_{p_i}$ from which we obtain that $F(\boldsymbol{B'})(p_i) = 1$ and hence that $F(\boldsymbol{B'}) \models \neg\chi_i$. By the anonymity and systematicity of F it therefore follows that $F(\boldsymbol{B'}) \models \bigwedge_{i\in\mathcal{N}}\neg\chi_i$ as by construction the individual positions on each issue are identical, modulo permutations. $\boldsymbol{B'}$ is therefore a truthful but totally inefficient ballot profile. Since $\boldsymbol{B'}$ is, by construction, a profile where each voter is truthful, by Proposition 4, $\boldsymbol{B'}$ is also a NE in weakly dominant strategies. This completes the proof. □

3.2.4 Equilibria in Uniform Aggregation Games

Since constant payoffs are special cases of uniform ones, negative results such as Proposition 8 still hold for uniform aggregation games. As to truthful voting, the positive result of Proposition 4 does not generalise to uniform aggregation games:

Proposition 9. *There exist uniform aggregation games for a non-manipulable aggregator in which truthful strategies are not dominant.*

Proof. Define the uniform aggregation game as follows. Let $\mathcal{I} = \{p, q, t\}$, $\mathcal{N} = \{1, 2, 3\}$ and $F = \mathsf{maj}$. Let $\gamma_1 = \neg p \wedge q \wedge \neg t$, $\gamma_2 = \neg p \wedge \neg q \wedge \neg t$, and $\gamma_3 = \neg p \wedge \neg q \wedge t$. Define the payoff functions as follows, let $\pi_i(B) = 1$ for $i = 3$ and $B = (0, 1, 0)$, and 0 otherwise. Take the following profiles: $\boldsymbol{B}_1 = ((0, 1, 0), (0, 0, 0), (0, 0, 1))$ and $\boldsymbol{B}_2 = ((0, 1, 0), (0, 0, 0), (0, 1, 0))$. Since $\mathsf{maj}(\boldsymbol{B}_1) = (0, 0, 0)$ and $\mathsf{maj}(\boldsymbol{B}_2) = (0, 1, 0)$, we have $\boldsymbol{B}_2 \succ_3^{\pi} \boldsymbol{B}_1$ and $\boldsymbol{B}_1$, unlike $\boldsymbol{B}_2$, contains a truthful strategy by 3. □

The fact that truthful voting is not always a dominant strategy for aggregation games with simple goals might seem counterintuitive, especially when the payoff is required to be uniform across the profiles that lead to the same outcome. The reason for this lies in the effect of the payoff function. When a player is in the position of changing the outcome of the decision in a certain profile, this does not necessarily imply she has the power to make the collective decision satisfy her goal. She may only be able to lead the group to a decision which, even though still not satisfying her goal, yields a better payoff for her.

Despite the negative result in Proposition 9, we can still prove the existence of truthful and efficient equilibria in a uniform aggregation game if we assume the mutual consistency of the individual goals of a resilient winning coalition.

Proposition 10. *Every C-consistent uniform aggregation game for non-manipulable F has a NE that is C-truthful and C-efficient, if $C \in \mathcal{W}^{+}_{\bigwedge\Gamma}$ where $\Gamma = \{\gamma_i \mid i \in C\}$.*

Proof. Take a C-consistent uniform aggregation game. Note that since each goal γ_i is a cube, also $\bigwedge\Gamma$ is a cube, so that Definition 3 applies. There exists then a ballot B^* such that $B^* \models \bigwedge\Gamma$. Take now any ballot profile $\boldsymbol{B}^*$ such that B^* is the ballot of all and only the voters in C while all agents in $\mathcal{N}\backslash C$ vote the inverse ballot $\overline{B^*}$ (that is, for any issue j, $B^*(j) = 1$ if and only if $\overline{B}^*(j) = 0$). Since $C \in \mathcal{W}_{\bigwedge\Gamma}$ (by the assumption that $C \in \mathcal{W}^{+}_{\bigwedge\Gamma}$) we have that $F(\boldsymbol{B}^*)$ satisfies the conjunction of the goals of the individuals in C, and each individual in C votes truthfully. We show that $\boldsymbol{B}^*$ is a NE, by showing that (a) no agent in $\mathcal{N}\backslash C$ has a profitable deviation, and (b) no agent in C has a profitable deviation. As to (a), since F is monotonic, any change in the ballot $\overline{B}^*(j)$ by some voter in $\mathcal{N}\backslash C$ does not change the outcome $F(\boldsymbol{B}^*) = B^*$. As to (b), any change in the ballot B^* by some voter in C does not change the outcome because $C \in \mathcal{W}^{+}_{\bigwedge\Gamma}$. □

3.2.5 Discussion

Let us recapitulate the findings of this section. We have shown, for a well-behaved class of aggregators—the non-manipulable ones—, that aggregation games with constant payoffs have many NE, since truthful ballots are weakly dominant strategies in such games (Proposition 4) and that, for the subclass of non-manipulable aggregators known as quota rules, truthful ballots are exactly the set of weakly dominant strategies (Proposition 6).

We then showed that such results do not carry over to uniform aggregation games (Propositions 9 and 8), but that in such games, however, the existence of truthful *and* efficient equilibria can be guaranteed whenever a resilient winning coalition has non-conflicting goals (Proposition 10).

Finally, Proposition 8 has highlighted a key issue of aggregation games: truthful equilibria may be totally inefficient in the sense of failing to satisfy the goals of all voters, even when all such goals are consistent. The purpose of the following section is to introduce an endogenous pre-play negotiation mechanism which, in equilibrium, allows players to select truthful and efficient NE of the underlying aggregation game, thus resolving the tension between truthfulness and efficiency.

4. Pre-vote Negotiations

This section presents endogenous aggregation games: aggregation games augmented with a pre-vote negotiation phase. In a nutshell voters will now be allowed, before the vote takes place, to sacrifice a part of their expected gains in order to influence the other voters' decision-making. We show that allowing such negotiations: (i) guarantees the selection of efficient equilibria for all individuals when one such equilibrium exists (ii) discards all equilibria that are inefficient for the voters of any winning coalition.

4.1 Endogenous Aggregation Games

Endogenous aggregation games consist of two phases:

- A *pre-vote phase*, where, starting from a uniform aggregation game, players make simultaneous transfers of utility that may modify each others' payoffs in the game;
- A *vote phase*, where players play the aggregation game resulting from the original game after payoffs are updated according to the tranfers made in the pre-vote phase.

As usual, it is assumed that the players have common knowledge of the structure of the game (including their goals and payoffs). A key assumption is furthermore that the transfers made in the pre-vote phase be binding, for instance through a central authority.[21] The authority would, besides running the election in the vote phase, also collect and enforce the transfers announced in the pre-vote phase.

Pre-vote strategies are modelled as **transfer functions** of the form:

$$\tau_i : \mathcal{D}^{\mathcal{N}} \times \mathcal{N} \to \mathbb{R}_+$$

where $i \in \mathcal{N}$. These functions encode the amount of payoff that player i commits to give to player j should a given ballot profile $\boldsymbol{B}$ be played, in symbols, $\tau_i(\boldsymbol{B}, j)$. The set of all transfer functions is denoted by $\mathcal{T}$, and a **transfer profile** is a tuple of transfer functions $\tau \in \mathcal{T}^{|\mathcal{N}|}$. We denote by τ^0 the 'void' transfer where at every profile every player gives 0 to the other players.

The aggregation game induced by the transfer profile τ from $\mathcal{A}$ is denoted $\tau(\mathcal{A}) = \langle \mathcal{N}, \mathcal{I}, F, \{\gamma_i\}_{i\in\mathcal{N}}, \{\tau(\pi)_i\}_{i\in\mathcal{N}} \rangle$ where, for any $i \in \mathcal{N}$:

$$\tau(\pi)_i(\boldsymbol{B}) = \pi_i(\boldsymbol{B}) + \sum_{j\in\mathcal{N}} (\tau_j(\boldsymbol{B}, i) - \tau_i(\boldsymbol{B}, j)) \tag{3}$$

So the payoff of player i at profile $\boldsymbol{B}$ *once* τ *is played*, consists of the old payoff that i was receiving at $\boldsymbol{B}$, *plus* the money that i receives from the other players at $\boldsymbol{B}$, *minus* what i gives to them at $\boldsymbol{B}$. Notice that transfers do not preserve the uniformity of payoffs: even though $\mathcal{A}$ is always assumed to be uniform, $\tau(\mathcal{A})$ is not necessarily so,[22] and may lack a NE (recall Fact 3).

It is important to notice that while our pre-vote phase is based on Jackson and Wilkie's (2005) endogenous transfer functions, their effect on the resulting games, and therefore the resulting equilibria, is fundamentally different. In particular: our voting games are based on lexicographic preferences and are therefore not reducible to strategic games (Turrini, 2016), which are instead the object of study of Jackson and Wilkie; our equilibrium analysis will moreover not rely on mixed strategies, which guarantee the existence of equilibria in the strategic games studied by Jackson and Wilkie, but are somewhat harder to interpret in a voting context.

Endogenous aggregation games are defined formally as follows:

21. The existence of such authority is a common assumption in work on election control and bribing (Faliszewski & Rothe, 2016), as well as persuasion (Hazon, Lin, & Kraus, 2013). It is often referred to as 'chair' or 'election organiser'.
22. In fact it is easy to see that there always exists a transfer that turns a uniform aggregation game into a non uniform one.

Definition 6 (Endogenous aggregation games)**.** *An endogenous aggregation game is a tuple* $\mathcal{A}^{\mathcal{T}} = \langle \mathcal{A}, \{T_i\}_{i \in N} \rangle$ *where* $\mathcal{A}$ *is a uniform aggregation game, and for each* $i \in \mathcal{N}$, $T_i = \mathcal{T}$. *A* constant *endogenous aggregation game is an endogenous aggregation game* $\mathcal{A}^{\mathcal{T}}$ *where* $\mathcal{A}$ *is assumed to be constant.*[23]

In an endogenous aggregation game, each player i selects first a transfer τ_i (pre-vote phase). The resulting transfer profile τ yields the new aggregation game $\tau(\mathcal{A})$. Then each player selects a ballot B_i (vote phase) in $\tau(\mathcal{A})$. The resulting ballot profile $\boldsymbol{B}$ yields the collective ballot $F(\boldsymbol{B})$, where F is the aggregator of $\mathcal{A}$.

4.2 Equilibrium Selection via Pre-vote Negotiations

In this section we first provide the definition of the solution concept we use for analysing endogenous games as two-stage extensive form games, and we then present our main results on the selection of efficient equilibria via pre-vote negotiations.

4.2.1 Solving Endogenous Aggregation Games

Endogenous aggregation games are (perfect information) extensive form games with two stages of simultaneous choices. So in an endogenous aggregation game $\mathcal{A}^{\mathcal{T}}$, a strategy of player i is a pair $\sigma_i \in \mathcal{T} \times \mathcal{D}^{\mathcal{T}^{\mathcal{N}}}$, that is, a choice of a transfer τ_i in the pre-vote phase (first element), followed by the choice of a ballot B_i in *each* aggregation game that results from some possible transfer profile (second element). A profile $\sigma = \langle \sigma_1, \ldots, \sigma_{|\mathcal{N}|} \rangle \in \left(\mathcal{T} \times \mathcal{D}^{\mathcal{T}^{\mathcal{N}}}\right)^{\mathcal{N}}$ has the following important characteristics. First of all, it selects a unique transfer profile and a unique ballot profile—**transfer-ballot profile** in short—$(\tau, \boldsymbol{B})$ that will be played if σ is chosen by the players. We refer to such transfer and ballot profiles as being **induced by** σ. Secondly, for any transfer profile τ, a strategy profile σ uniquely determines a ballot profile, which we denote $\sigma(\tau)$. We refer to such ballot profile as the ballot **induced by** σ **after** τ is played. The latter aspect of strategy profiles in endogenous aggregation games is worth stressing. Each strategy profile does not only determine a specific choice of transfers and ballots, but also a choice of ballots that would be played if other transfers were to be chosen, i.e., the counterfactual choices that would be made if other transfers were chosen by the players in the pre-vote phase of the game.

Given the above, it would seem natural to resort to (pure strategy) subgame perfect Nash equilibrium (SPE) to solve endogenous aggregation games. There is a complication however. As we observed earlier, the aggregation game resulting from a transfer profile may not be uniform and, therefore, may not necessarily have a NE. This makes SPE inapplicable as some subgames of the initial extensive game may, therefore, be unsolvable. Also notice how resorting to mixed strategy equilibria would not help addressing this issue, as lexicographic preferences are well-known for not being representable in terms of utility functions, therefore making Nash's NE existence theorem unavailable (Rubinstein, 2012).

What we do instead is to analyse endogenous aggregation games by assuming that, if the game resulting from a transfer profile does not have a NE, players play a maxmin strategy and resort to their maxmin value, or *security level*,[24] to evaluate the resulting aggregation

23. Constant endogenous aggregation games will be discussed in Section 5.

24. See Shoham and Leyton-Brown (2008) for an extensive exposition of the notion.

game.[25] In our quasi-dichotomous setup a maxmin strategy (ballot) is defined as:

$$\mathsf{argmax}_{B_i}^{\succeq_i^\pi} \mathsf{min}_{\boldsymbol{B}_{-i}}^{\succeq_i^\pi} \{F(\boldsymbol{B}) \mid \boldsymbol{B} = (B_i, \boldsymbol{B}_{-i})\}, \tag{4}$$

i.e., a (not necessarily unique) ballot that maximises the minimum, with respect to $\succeq_i^\pi$ (Definition 2), outcome for i. If the maxmin strategy B_i guarantees that γ_i is satisfied no matter what the other players do, then we say that B_i is **safe for** γ_i. The minimum payoff guaranteed by i's maxmin strategies is i's **security level**.

We can now move to the definition of the solution for an endogenous aggregation game. To do that, however, we first need the following auxiliary definition:

Definition 7 (Transfer game)**.** *Let $\mathcal{A}^{\mathcal{T}}$ be an endogenous aggregation game and fix a strategy profile σ. The* transfer game induced by σ *is the game in normal form $\mathcal{A}^\sigma = \langle \mathcal{N}, \{T_i\}_{i\in\mathcal{N}}, \{\succeq_i^\sigma\}_{i\in\mathcal{N}} \rangle$ where $T_i = \mathcal{T}$ and each $\succeq_i^\sigma \subseteq \mathcal{T}^2$ is the preference over transfer profiles defined as follows. Given a strategy profile σ, a transfer profile τ is said to satisfy goal γ_i (in symbols, $\tau \models \gamma_i$) if* either *the ballot profile $\sigma(\tau) = (B_1, \ldots, B_{|\mathcal{N}|})$ induced by σ after τ is a NE of $\tau(\mathcal{A})$ and $F(\sigma(\tau)) \models \gamma_i$,* or *$B_i$ is safe for γ_i in $\tau(\mathcal{A})$, and B_i is a maxmin strategy of i. The payoff π_i of profile τ is* either *$\tau(\pi)_i(\sigma(\tau))$ if $\sigma(\tau)$ is a NE of $\mathcal{A}^{\mathcal{T}}$,* or *$i$'s security level in $\tau(\mathcal{A})$ otherwise. Given the above, for any $i \in \mathcal{N}$, $\tau \succ_i^\sigma \tau'$ if and only if any of the two following conditions hold:*[26]

i) $F(\sigma(\tau')) \not\models \gamma_i$ and $F(\sigma(\tau)) \models \gamma_i$;

ii) $F(\sigma(\tau')) \models \gamma_i$ if and only if $F(\sigma(\tau)) \models \gamma_i$, and $\pi_i(\sigma(\tau)) > \pi_i(\sigma(\tau'))$.

The two transfer profiles are equally preferred whenever $F(\sigma(\tau')) \models \gamma_i$ if and only if $F(\sigma(\tau)) \models \gamma_i$, and $\pi_i(\sigma(\tau)) = \pi_i(\sigma(\tau'))$. The resulting weak preference order among trasfer profiles is denoted $\succeq_i^\sigma$.

Observe that the preferences $\succeq_i^\sigma$ over transfer profiles are therefore quasi-dichotomous. Intuitively, a transfer game is the game that a σ needs to solve in the negotiation phase of an endogenous aggregation game, once the continuations $\sigma(\tau)$ have been fixed for any transfer profile τ. We are now in the position to state formally our definition of a solution concept for endogenous games:

Definition 8 (Solutions)**.** *Let $\mathcal{A}^{\mathcal{T}}$ be an endogenous aggregation game. A strategy profile $\sigma \in \left(\mathcal{T} \times \mathcal{D}^{\mathcal{T}^{\mathcal{N}}}\right)^{\mathcal{N}}$ of $\mathcal{A}^{\mathcal{T}}$ is a* solution *of $\mathcal{A}^{\mathcal{T}}$ if and only if:*

(1) Voting phase. *The ballot profile $\sigma(\tau) \in \mathcal{D}^{\mathcal{N}}$ induced by σ after τ is a NE of the aggregation game $\tau(\mathcal{A})$, if such equilibrium exists, or a profile of maxmin ballots of $\tau(\mathcal{A})$, otherwise.*

25. A game with no NE should be seen as an unstable game, i.e., a situation in which the players do not have any reason to believe that some specific outcome will be realised. Measures such as the security level therefore compensate for this uncertainty. We could have adopted a number of alternative solutions, e.g., taking the value of the minimum outcome for each player, or considering such games never to be profitable deviations. Ultimately, all these assumption have the effect of ruling out profitable deviations to games with no NE.

26. Cf. earlier Definition 2.

(2) Negotiation phase. *The first element in pair σ is a transfer profile τ, which is a NE of the transfer game $\mathcal{A}^{\sigma}$.*

We refer to the transfer-ballot profile induced by a solution σ as the **solution outcome** of $\mathcal{A}^{\mathcal{T}}$. So, intuitively, a solution outcome is obtained by constructing a profile σ through a backward induction procedure that starts with the voting phase, and then moves to the negotiation phase:

- *Voting phase* In the aggregation game resulting after each transfer profile, a NE is selected where at least one such equilibrium exists, or a maxmin profile is selected otherwise; the value of such NE, if it exists, or the players' security levels, otherwise, are used to evaluate transfer profiles in the next phase (negotiation phase) of the procedure.
- *Negotiation phase* A transfer profile is selected, such that no profitable deviation exists to another transfer profile, for any player, given the continuations that were selected in the first phase (the voting phase) of the procedure.

4.2.2 Surviving Equilibria

This section provides existence results for solutions of endogenous games. We actually provide stronger results, showing not only that solutions exist, but also (in the next sections) necessary and sufficient conditions for them to enjoy desirable properties in terms of the 'quality' of the equilibria that negotiation enables. In other words, we study endogenous aggregation games as endogenous mechanisms for the refinement of the equilibria of their underlying aggregation games.

Definition 9. *Let $\mathcal{A}$ be a uniform aggregation game, and $\boldsymbol{B}$ a NE of $\mathcal{A}$. $\boldsymbol{B}$ is a* surviving Nash equilibrium *(SNE) of $\mathcal{A}$ if there exists a transfer profile τ such that $(\tau, \boldsymbol{B})$ is a solution outcome of the endogenous aggregation game $\mathcal{A}^{\mathcal{T}}$.*

Intuitively, SNE identify those voting outcomes that can be rationally sustained by an appropriate pre-vote negotiation. Clearly, not all NE of the initial game will be surviving equilibria but, crucially, it can be shown that NE with good properties are of this kind, as we set out to show in this section. Let us first establish the following lemma.

Lemma 11. *Let $\mathcal{A}$ be a uniform aggregation game for a non-manipulable aggregator F. Then, for every $\mathcal{N}$-efficient and $\mathcal{N}$-truthful ballot profile $\boldsymbol{B}$ of $\mathcal{A}$, there exists a transfer profile τ such that $\boldsymbol{B}$ is a weakly dominant strategy equilibrium in $\tau(\mathcal{A})$.*

Proof. Let $\boldsymbol{B}$ be an $\mathcal{N}$-efficient and $\mathcal{N}$-truthful ballot profile of $\mathcal{A}$. We construct a transfer profile τ such that for any $i \in \mathcal{N}$ the strategy B_i is a strictly dominant strategy for i in $\tau(\mathcal{A})$, that is: for any profile $\boldsymbol{B}'$, $(B_i, \boldsymbol{B}'_{-i}) \succeq_i^{\pi} \boldsymbol{B}'$ and for some $\boldsymbol{B}'$, $(B_i, \boldsymbol{B}'_{-1}) \succ_i^{\pi} \boldsymbol{B}'$. The construction goes as follows.[27] Consider now the quantity:

$$M = 1 + \max\left(\left\{z \mid \exists \boldsymbol{B}, \boldsymbol{B}' \textit{and } i \in \mathcal{N} s.t.\ z = \pi_i(\boldsymbol{B}) - \pi_i(\boldsymbol{B}')\right\}\right) \tag{5}$$

27. Our construction is an adaptation of the one in Jackson and Wilkie (2005), Theorem 4. We however point out once more how our proofs do not rely on the existence of NE in each aggregation game.

that is, M exceeds by one unit the maximal difference between the payoff received at any two outcomes by any agent. We are now ready to define a transfer function. For all $i, j \in \mathcal{N}$:

$$\tau_i(\boldsymbol{B}', j) = \begin{cases} 2M & \text{if } B'_i \neq B_i \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

In words, each player i commits to pay each other player the sum $2M$ in case he deviates from the ballot B_i. By the definition of the preference relation $\succeq_i^{\pi}$ (Definition 2), in order to prove the claim we need to show that: $\boxed{\text{A}}$ For any $\boldsymbol{B}'$ if $F(\boldsymbol{B}') \models \gamma_i$ then $F(B_i, \boldsymbol{B}'_{-i}) \models \gamma_i$, that is, under no circumstance i's goal would become falsified by playing B_i. $\boxed{\text{B}}$ For any $\boldsymbol{B}'$ if $F(\boldsymbol{B}') \models \gamma_i$ if and only if $F(B_i, \boldsymbol{B}'_{-i}) \models \gamma_i$ (that is both profiles satisfy i's goals), then $\tau(\pi)_i(B_i, \boldsymbol{B}'_{-i}) \geq \tau(\pi)_i(\boldsymbol{B}')$, and for some $\boldsymbol{B}'$, $\tau(\pi)_i(B_i, \boldsymbol{B}'_{-i}) > \tau(\pi)_i(\boldsymbol{B}')$.

$\boxed{\text{Claim A}}$ Assume that $F(\boldsymbol{B}') \models \gamma_i$. There are two cases. First, $B'_i \models \gamma_i$. Since $\boldsymbol{B}$ is $\mathcal{N}$-truthful we know that $B_i \models \gamma_i$. Now γ_i is assumed to be a cube (Definition 1) so B'_i and B_i can differ only on the evaluation of issues on which the truth of γ_i does not depend. By the independence of F we conclude that $F(B_i, \boldsymbol{B}'_{-i}) \models \gamma_i$. Second, $B' \not\models \gamma_i$. But, as above, $B_i \models \gamma_i$. By the assumption $F(\boldsymbol{B}') \models \gamma_i$, so in profile $(B_i, \boldsymbol{B}'_{-i})$ voter i is changing the evaluation of at least one of the atoms that was falsifying γ_i in B'_i (recall that γ_i is a cube) to the evaluation given by $F(\boldsymbol{B}')$. By the monotonicity of F we can therefore conclude that $F(B_i, \boldsymbol{B}'_{-i}) \models \gamma_i$.

$\boxed{\text{Claim B}}$ Given the definition of payoffs in endogenous games (3), the claim follows directly from the construction of τ. This completes the proof. □

Intuitively, the lemma establishes that whenever a truthful profile exists that satisfies all players' goals, that profile can be turned into a dominant strategy equilibrium by means of a suitable combination of pre-vote transfers which, in essence, make the players' commitment to that equilibrium credible.

We finally give two examples of what happens if the assumptions behind the lemma are not satisfied. First, we describe a uniform aggregation game $\mathcal{A}$ for a non-manipulable F, and ballot profiles $\boldsymbol{B}$ that are either not $\mathcal{N}$-efficient or not $\mathcal{N}$-truthful, for which there exists no transfer profile τ such that $\boldsymbol{B}$ is a weakly dominant strategy equilibrium in $\tau(\mathcal{A})$.

Example 3. *Let $\mathcal{A}$ be such that $\mathcal{N} = \{1, 2, 3\}$, $\mathcal{I} = \{p\}$, $\gamma_i = p$, for each $i \in \mathcal{N}$, and let $F = \mathsf{maj}$. Let π be arbitrary. Consider now any profile $\boldsymbol{B}$ where for at least one agent i, $B_i(p) \neq 1$. Each such profile is not $\mathcal{N}$-truthful, and note that all profiles that are not $\mathcal{N}$-efficient are among those. It is easy to see that, no matter the utility function π, there is no transfer τ which would make the profile a weakly dominant strategy equilibrium of $\tau(\mathcal{A})$.*

Second, we show a uniform aggregation game $\mathcal{A}$ for an aggregator that is manipulable—specifically, that is not monotonic—, and an $\mathcal{N}$-efficient and $\mathcal{N}$-truthful ballot profile $\boldsymbol{B}$ for which there exists no transfer profile τ such that $\boldsymbol{B}$ is a weakly dominant strategy equilibrium in $\tau(\mathcal{A})$.

Example 4. *Let $\mathcal{A}$ be such that $\mathcal{N} = \{1, 2, 3\}$, $\mathcal{I} = \{p\}$, $\gamma_i = p$, for each $i \in \mathcal{N}$ and let π be arbitrary. Let F be such that $F(\boldsymbol{B})(p) = 1$ whenever $B_i(p) = 1$ for all $i \in \mathcal{N}$, or whenever $B_i(p) = 0$, for all $i \in \mathcal{N}$, and let $F(\boldsymbol{B}) = \mathsf{maj}(\boldsymbol{B})$, otherwise. Observe that in such a*

game, for no agent a truthful ballot is also a weakly dominant strategy. Independently of the choice of π, there is no transfer profile τ such that the $\mathcal{N}$-efficient and $\mathcal{N}$-truthful profile $\boldsymbol{B} = (1, 1, 1)$ becomes a weakly dominant strategy equilibrium in $\tau(\mathcal{A})$.

4.2.3 Good Equilibria Survive

Lemma 11 establishes the existence of a suitable transfer profile sustaining 'good'—that is, truthful and efficient—equilibria. We now show that endogenous aggregation games have solutions which require precisely the transfer profile constructed in the proof of Lemma 11 to be played in the negotiation phase of the game, and which therefore lead the voters to play 'good' equilibria in the voting phase.

Theorem 12. *Let $\mathcal{A}$ be a uniform aggregation game for a non-manipulable aggregator F. Every $\mathcal{N}$-efficient and $\mathcal{N}$-truthful NE of $\mathcal{A}$ is a SNE.*

Proof. Let $\boldsymbol{B}$ be a $\mathcal{N}$-efficient and $\mathcal{N}$-truthful NE of $\mathcal{A}$. By the assumptions on the aggregator and Lemma 11, we can construct a transfer profile τ according to formula (6) such that $\boldsymbol{B}$ is a weakly dominant strategy equilibrium of $\tau(\mathcal{A})$. We have to show that $(\tau, \boldsymbol{B})$ is a solution outcome of $\mathcal{A}^{\mathcal{T}}$. Suppose towards a contradiction that this is not the case, that is, that there exists a profitable deviation by player i from the strategy profile that induces the transfer-ballot profile $(\tau, \boldsymbol{B})$. Since $\boldsymbol{B}$ is a dominant strategy equilibrium in $\tau(\mathcal{A})$ such deviation has to involve a different transfer τ_i^* by player i in the pre-vote phase. We identify two cases:

Case 1 The deviation induces a new transfer profile $\tau' = (\tau_i^*, \tau_{-i})$ such that $\tau'(\mathcal{A})$ has no NE. By Definition 7 a deviation to τ_i^* would yield i his security level in $\tau'(\mathcal{A})$. We show that $\tau \succeq_i^{\pi} \tau'$, and hence the deviation cannot be profitable for i. Since $\boldsymbol{B} \models \gamma_i$ we have, by Definition 8, that if $\tau' \models \gamma_i$ then $\tau \models \gamma_i$, that is, it cannot be the case that τ' leads i to satisfy its goal while τ does not. As to the payoffs, observe that $\pi_i(\tau)$ is the payoff yielded by the NE $\boldsymbol{B}$ of $\tau(\mathcal{A})$, and that given the construction of τ' no agent transfers any utility to i in τ': for all $j \in \mathcal{N} \setminus \{i\}$, $\tau'_j(\boldsymbol{B}', i) = 0$ whenever $B'_j = B_j$. It follows that for every ballot profile $\boldsymbol{B}'$ in $\tau'(\mathcal{A})$ i's payoff is at most as much as what it obtains from $\boldsymbol{B}$ in $\tau(\mathcal{A})$. Hence, $\tau \succeq_i^{\pi} \tau'$.

Case 2 The deviation induces a new transfer profile $\tau' = (\tau_i^*, \tau_{-i})$ such that the game $\tau'(\mathcal{A})$ has a NE. We show that τ_i^* cannot be a profitable deviation for i. Since the strategy profile inducing $(\tau, \boldsymbol{B})$ is assumed to be a solution, i's deviation determines a transfer-ballot profile $(\tau', \boldsymbol{B}')$ where $\boldsymbol{B}'$ may or may not be a NE. However, if i's deviation toward $(\tau', \boldsymbol{B}')$ where $\boldsymbol{B}'$ is not a NE is profitable, i must have another at least as profitable deviation determining the transfer-ballot profile $(\tau', (B_i^*, \boldsymbol{B}'_{-i}))$, where $(B_i^*, \boldsymbol{B}'_{-i})$ is a NE of $\tau'(\mathcal{A})$. This, notice, has to be the case, as $\tau'(\mathcal{A})$ has a NE and thus the other players are necessarily playing in equilibrium. We therefore assume w.l.o.g., that the transfer-ballot profile $(\tau', \boldsymbol{B}')$ determined by i's deviation is such that $\boldsymbol{B}'$ is a NE of $\tau(\mathcal{A})$. Observe now that in order for $(\tau', \boldsymbol{B}')$ to be profitable for i there must exist a $j \neq i$ such that $B'_j \neq B_j$, that is, there exists at least another agent j besides i who contributes to deviating from $\boldsymbol{B}$ to $\boldsymbol{B}'$. This is because $\boldsymbol{B}$ is an $\mathcal{N}$ efficient NE and by construction of τ',we have that $\tau'_j(\boldsymbol{B}', i) = 0$ whenever $B'_j = B_j$. So let there be $k \geq 1$ players $j \neq i$ for which $B'_j \neq B$ and consider some such j. We have two subcases:

Case 2a $\boldsymbol{B'} \not\models \gamma_j$. Note that since $\boldsymbol{B'}$ is a NE, we have that no B''_j is such that $F(B''_j, \boldsymbol{B'}_{-j}) \models \gamma_j$, so B'_j is a best response only by virtue of j's payoff. By the construction of τ', playing B'_j gives j the following payoff:

$$\tau'(\pi)_j(\boldsymbol{B'}) = \pi_j(\boldsymbol{B'}) - (|\mathcal{N}| - 1)2M + 2M(k-1) + \tau'_i(\boldsymbol{B'}, j).$$

If j plays B_j instead, then j's payoff is:

$$\tau'(\pi)_j(B_j, \boldsymbol{B'}_{-j}) = \pi_j(B_j, \boldsymbol{B'}_{-j}) + 2M(k-1) + \tau'_i(B_j, \boldsymbol{B'}_{-j}, j).$$

Now since $\boldsymbol{B'}$ is a NE by assumption and j does not have a better response that can satisfy her goal, we have that $\tau'(\pi)_j(\boldsymbol{B'}) \geq \tau'(\pi)_j(B_j, \boldsymbol{B'}_{-j})$ and therefore:

$$\tau'_i(\boldsymbol{B'}, j) - \tau'_i(B_j, \boldsymbol{B'}_{-j}, j) \geq \pi_j(B_j, \boldsymbol{B'}_{-j}) - \pi_j(\boldsymbol{B'}) + (|\mathcal{N}| - 1)2M.$$

Now we use this inequality to compare i's payoff in $(\tau, \boldsymbol{B})$ vs. $(\tau', \boldsymbol{B'})$. Given the definition of M in Formula (5) and given the fact that $|\mathcal{N}| - 1 \geq 2$ it follows that $\tau'_i(\boldsymbol{B'}, j) - \tau'_i(B_j, \boldsymbol{B'}_{-j}, j) \geq 3M$, which in turn implies that $\tau'_i(\boldsymbol{B'}, j) \geq 3M$. Therefore i's payoff $\tau'(\pi)_i(\boldsymbol{B'})$ in the new NE $\boldsymbol{B'}$ is at most:

$$\pi_i(\boldsymbol{B'}) - k3M + k2M.$$

In contrast, by the construction of τ in formula (6), $\tau(\pi)_i(\boldsymbol{B}) = \pi_i(\boldsymbol{B})$. So from the fact that $k \geq 1$ it follows that $\pi_i(\boldsymbol{B'}) - k3M + k2M \leq \pi_i(\boldsymbol{B})$ and therefore that $\tau'(\pi)_i(\boldsymbol{B'}) \leq \tau(\pi)_i(\boldsymbol{B})$. Since $\boldsymbol{B}$ is $\mathcal{N}$-efficient, by Definition 8, $\tau \succeq_i^{\pi} \tau'$ and the constructed deviation τ_i^* cannot be profitable.

Case 2b $F(\boldsymbol{B'}) \models \gamma_j$. Notice that from this it follows that $F(B_j, \boldsymbol{B'}_{-j}) \models \gamma_j$, for otherwise there would be a ballot profile (i.e., $(B_j, \boldsymbol{B'}_{-j})$) where B'_j is a better response for j making γ_j true. From this we would conclude that $\boldsymbol{B}$ in $\tau(\mathcal{A})$ would not be a dominant strategy equilibrium, against our assumption. We can then proceed as in Case 2a, showing that $\tau'(\pi)_i(\boldsymbol{B'}) \leq \tau(\pi)_i(\boldsymbol{B})$. Since, as just noticed, both $\boldsymbol{B}$ and $\boldsymbol{B'}$ satisfy γ_j, by Definition 8, $\tau \succeq_i^{\pi} \tau'$ and the constructed deviation τ_i^* cannot be profitable. This completes the proof. □

By combining Theorem 12 with Proposition 10, we also get the following corollary:

Corollary 13. *Let F be a non-manipulable aggregator. Every $\mathcal{N}$-consistent uniform aggregation game for F, where $\mathcal{N} \in \mathcal{W}^+_{\bigwedge \Gamma}$ with $\Gamma = \{\gamma_i \mid i \in \mathcal{N}\}$, has a SNE that is $\mathcal{N}$-truthful and $\mathcal{N}$-efficient.*

That is, in constant aggregation games where the grand coalition is resilient, there exists a SNE in which every agent votes truthfully and all goals are satisfied.

Finally, we show that the assumption of non-manipulability of the aggregator F in Theorem 12 is necessary.

Proposition 14. *There exist uniform aggregation games with F manipulable, for which there exists an $\mathcal{N}$-efficient and $\mathcal{N}$-truthful NE that is not a SNE.*

Proof. The proof is by construction. Let $\mathcal{A}$ be a uniform aggregation game with $\mathcal{N} = \{1,2,3\}$, $\mathcal{I} = \{1,2\}$, $\gamma_i = p_1$, for each $i \in \mathcal{N}$. Let moreover, for each $i \in \mathcal{N}$, $\pi_i(\boldsymbol{B}) = 1$ whenever $F(\boldsymbol{B}) = (1,1)$, and $\pi_i(\boldsymbol{B}) = 0$ otherwise. Let now F be such that $F(\boldsymbol{B}) = (1,1)$ whenever $B_i = (0,0)$, for each $i \in \mathcal{N}$, and $F(\boldsymbol{B}) = (\mathsf{maj}(\boldsymbol{B})(1), 0)$, otherwise. That is, F always rejects issue p_2 except when everybody rejects it, and accepts p_1 whenever either there is a majority accepting it or everybody rejects it. This game is clearly uniform and goals are cubes, F is independent but it is not monotonic and is therefore manipulable.

Consider now the profile $\boldsymbol{B}^*$ such that $B_i^* = (1,1)$, for each $i \in \mathcal{N}$. This profile is an $\mathcal{N}$-efficient and $\mathcal{N}$-truthful NE of $\mathcal{A}$. However it is not a SNE. To see this, we proceed towards a contradiction and assume that $\boldsymbol{B}^*$ is a SNE. So there exists a solution outcome $(\tau^*, \boldsymbol{B}^*)$ of the endogenous aggregation game. By the way we defined π, for all j we have that $\pi_j(\boldsymbol{B}^*) < \pi_i(\boldsymbol{B'})$, where $\boldsymbol{B'}$ is the profile at which $B'_j = (0,0)$, for each $j \in \mathcal{N}$. By the definition of transfer function, for any transfer function τ and any profile $\boldsymbol{B}$, $\sum_{j\in\mathcal{N}} \pi_j(\boldsymbol{B}) = \sum_{j\in\mathcal{N}} \pi_j^{\tau}(\boldsymbol{B})$. From these two facts we conclude that for some player i we have that $\pi_i^{\tau^*}(\boldsymbol{B}^*) < \pi_i^{\tau^*}(\boldsymbol{B'})$. Observe that in $\boldsymbol{B'}$ i's goal is still satisfied. We show that i can deviate from τ^* in such a way that the resulting voting game after such deviation, i) $\boldsymbol{B'}$ is a NE, and ii) it is unique. This would be a profitable deviation for i thereby showing that $(\tau^*, \boldsymbol{B}^*)$ cannot be a solution outcome. So we show how to construct a deviation from τ^* which satisfies the two above constraints.

i) To make sure that $\boldsymbol{B'}$ remains a NE in the voting game after the new transfer profile it suffices for the deviation τ_i of i to be such that $(\tau_i, \tau^*_{-i})(\boldsymbol{B'}) = \tau^*(\boldsymbol{B'})$. That is payoffs for $\boldsymbol{B'}$ remain as in the original game $\mathcal{A}$.

ii) Now we have to show how the deviation τ_i of i can be constructed so that no other profile exists which is a NE in $(\tau_i, \tau^*_{-i})(\mathcal{A})$ and which satisfies i's goal (as otherwise the deviation we are trying to construct would not be profitable for i). To do this we set i's transfer deviation to guarantee that, in any profile that satisfies his goal, but where any agent's ballot is different from $(0,0)$, some agent has an incentive to deviate from his current ballot to another one, that is, has a better response. Such deviation τ_i can be built as follows. W.l.o.g., below we assume $i = 1$.[28]

(1) Profiles in which p_1 is satisfied and 1 is not pivotal to p_1 being satisfied (i.e., 2 and 3 both accept p_1). Transfer strategy τ_1 should be set so that: $(\tau_1, \tau^*_{-1})(\boldsymbol{B'}) = \tau^*(\boldsymbol{B'})$; and for any $\boldsymbol{B''}$ and $\boldsymbol{B'''}$, $\pi_1^{(\tau_1,\tau^*_{-1})}(\boldsymbol{B''}) > \pi_1^{(\tau_1,\tau^*_{-1})}(\boldsymbol{B'''})$ whenever, by interpreting the ballot profiles as two-bits binary numbers, $B''_1 < B'''_1$, that is according to the ordering, $(0,0) < (0,1) < (1,0) < (1,1)$. Observe that such transfer strategy commits 1 to always play $(0,0)$ whenever his goal is satisfied and he is not pivotal for the goal to be satisfied.

28. The construction we propose bears similarities with the concept of *hard equilibrium* induction studied in Harrenstein, Turrini, and Wooldridge (2014, 2017).

(2) Profiles in which p_1 is satisfied, 1 is not pivotal and $B_1 = (0,0)$. Transfer strategy τ_1 should be set so that:

$$\pi_2^{(\tau_1,\tau_{-1}^*)}((0,0),(1,1),(1,1)) < \pi_2^{(\tau_1,\tau_{-1}^*)}((0,0),(1,0),(1,1))$$
$$\pi_3^{(\tau_1,\tau_{-1}^*)}((0,0),(1,1),(1,0)) < \pi_3^{(\tau_1,\tau_{-1}^*)}((0,0),(1,1),(1,1))$$
$$\pi_2^{(\tau_1,\tau_{-1}^*)}((0,0),(1,0),(1,0)) < \pi_2^{(\tau_1,\tau_{-1}^*)}((0,0),(1,1),(1,0))$$
$$\pi_3^{(\tau_1,\tau_{-1}^*)}((0,0),(1,0),(1,1)) < \pi_3^{(\tau_1,\tau_{-1}^*)}((0,0),(1,0),(1,0))$$

Observe that such transfer strategy guarantees that whenever 1 plays $(0,0)$ and he is not pivotal for his goal to be satisfied, 2 and 3 always have better responses (preserving the satisfaction of their goal, but guaranteeing a higher payoff) to voting strategies which are different from $(0,0)$.

(3) Profiles in which p_1 is not satisfied, no player is pivotal and $B_1 = (0,0)$. Transfer strategy τ_1 should be set so that:

$$\pi_3^{(\tau_1,\tau_{-1}^*)}((0,0),(0,1),(0,1)) < \pi_3^{(\tau_1,\tau_{-1}^*)}((0,0),(0,1),(0,0))$$
$$\pi_2^{(\tau_1,\tau_{-1}^*)}((0,0),(0,1),(0,0)) < \pi_2^{(\tau_1,\tau_{-1}^*)}((0,0),(0,0),(0,0))$$
$$\pi_3^{(\tau_1,\tau_{-1}^*)}((0,0),(0,0),(0,1)) < \pi_3^{(\tau_1,\tau_{-1}^*)}((0,0),(0,0),(0,0))$$

Observe again that such transfer strategy guarantees that whenever 1 plays $(0,0)$ and no player is pivotal for his goal to be satisfied, 2 and 3 always have better responses (not reaching the goal, but guaranteeing a higher payoff) to voting strategies which are different from $(0,0)$.

(4) Profiles in which p_1 is satisfied, 1 is pivotal and $B_1 = (1,0)$. Transfer strategy τ_1 should be set so that:

$$\pi_2^{(\tau_1,\tau_{-1}^*)}((1,0),(0,0),(1,0)) < \pi_2^{(\tau_1,\tau_{-1}^*)}((1,0),(0,1),(1,0))$$
$$\pi_3^{(\tau_1,\tau_{-1}^*)}((1,0),(0,1),(1,0)) < \pi_3^{(\tau_1,\tau_{-1}^*)}((1,0),(0,1),(1,1))$$
$$\pi_2^{(\tau_1,\tau_{-1}^*)}((1,0),(0,1),(1,1)) < \pi_2^{(\tau_1,\tau_{-1}^*)}((1,0),(1,1),(1,1))$$
$$\pi_3^{(\tau_1,\tau_{-1}^*)}((1,0),(1,0),(0,0)) < \pi_3^{(\tau_1,\tau_{-1}^*)}((1,0),(1,0),(0,1))$$
$$\pi_2^{(\tau_1,\tau_{-1}^*)}((1,0),(1,0),(0,1)) < \pi_2^{(\tau_1,\tau_{-1}^*)}((1,0),(1,1),(0,1))$$
$$\pi_3^{(\tau_1,\tau_{-1}^*)}((1,0),(1,1),(0,1)) < \pi_3^{(\tau_1,\tau_{-1}^*)}((1,0),(1,1),(1,1))$$

Observe that such transfer strategy guarantees that whenever 1 plays $(1,0)$ and he is pivotal for his goal to be satisfied, 2 and 3 always have better responses (preserving the satisfaction of their goal, but guaranteeing a higher payoff) to voting strategies which are different from $(0,0)$.

It should be clear that the above four cases cover all profiles of relevance for the construction, and that constructing a transfer that meets the above constraints is easy: since the goals of all players remain unchanged throughout, and players can always transfer enough incentives, it suffices for i to transfer ϵ to the agent he wants to sway from her current strategy. After (τ_1, τ^*_{-1}), $\boldsymbol{B'}$ is therefore the only NE of the resulting game. The transfer strategy τ_i is therefore a profitable deviation for i, from which follows that $\boldsymbol{B}^*$ is not a SNE. Contradiction. □

4.2.4 Only Good Equilibria Survive

We now turn to necessary conditions for equilibria to survive, and observe that for an equilibrium to survive, it has to be efficient. This can be regarded as a converse statement of Theorem 12, and it holds true in a general form concerning winning coalitions (and not just the coalition $\mathcal{N}$) with internally consistent goals:

Theorem 15. *Let $\mathcal{A}$ be a C-consistent uniform aggregation game for a non-manipulable aggregator F, and such that $C \in \mathcal{W}_{\bigwedge \Gamma}$ where $\Gamma = \{\gamma_i \mid i \in C\}$. Then, every SNE of $\mathcal{A}$ is C-efficient.*

Proof. We proceed by contraposition. Let $\boldsymbol{B}$ be a NE that is not C-efficient, i.e., such that $F(\boldsymbol{B}) \not\models \gamma_i$ for some individual $i \in C$, and assume towards a contradiction that $\boldsymbol{B}$ is a SNE. Therefore there exists a transfer profile τ such that $(\tau, \boldsymbol{B})$ is a solution outcome of the endogenous game $\mathcal{A}^{\mathcal{T}}$. We proceed towards a contradiction and construct a profitable deviation for a player i from τ, that is a τ' such that $\tau' \succ_i^{\pi} \tau$. By C-consistency of $\mathcal{A}$ there exists a ballot B' such that $B' \models \bigwedge \Gamma$, hence in particular $B' \models \gamma_i$. Let now i deviate to any transfer profile $\tau' = (\tau'_i, \tau_{-i})$ such that she offers to all other players in C more than their payoff difference if they switch to vote for ballot B' while everybody else in C does so, i.e.,

$$\tau'_i((\boldsymbol{B}'_C, \boldsymbol{B}''_{-C}), j) > \tau(\pi)_j(\boldsymbol{B}'_{C-\{j\}}, B''_j, \boldsymbol{B}''_{-C}) - \tau(\pi)_j(\boldsymbol{B}'_C, \boldsymbol{B}''_{-C})$$

for each $j \in C$, each B''_j, and each $\boldsymbol{B}''_{-C}$, and where $\boldsymbol{B}'_C = (B'_j)_{j \in C}$. By the fact that B' is C-efficient, F independent and monotonic, and C is a winning coalition for $\bigwedge \Gamma$, this transfer makes each B'_j, with $j \in C$ a best response in profiles $(\boldsymbol{B}'_C, \boldsymbol{B}''_{-C})$, for any $\boldsymbol{B}''_{-C}$. It follows that each $(\boldsymbol{B}'_C, \boldsymbol{B}''_{-C})$ is a NE of $\tau'(\mathcal{A})$, which satisfies $\bigwedge \Gamma$, while $\boldsymbol{B}$ in $\tau(\mathcal{A})$ does not. We can therefore conclude (by Definition 8) that $\tau' \succ_i^{\pi} \tau$, completing the proof. □

The proof of Theorem 12 provides every agent with a simple algorithm to compute a negotiation strategy that will guarantee the emergence of an efficient equilibrium in the resulting game. Note that the use of cubes sidesteps the intractability of the satisfiability problem for the conjunction of the goals.

Observe also that Theorem 15 implies the existence of uniform aggregation games where no equilibrium is surviving, which in turn implies that Theorem 12 cannot be weakened from $\mathcal{N}$-efficiency to C-efficiency. This is the case when distinct but overlapping coalitions have incompatible goals, as the following example shows:

Example 5. *Let there be five players in $\mathcal{N}$, and let F be the majority rule. Let $\gamma_1 = p \wedge \neg r$, $\gamma_2 = \gamma_3 = \gamma_4 = \top$ and let $\gamma_5 = r \wedge \neg p$. Both coalitions $C_1 = \{1, 2, 3, 4\}$ and $C_2 = \{2, 3, 4, 5\}$ are resilient winning coalitions, and the game is both C_1-consistent and*

C_2-consistent. Hence, by Theorem 15 any surviving equilibria must be both C_1-efficient and C_2-efficient, which is impossible given that the conjunction of the goals of the two coalitions is inconsistent.

4.2.5 Summary

Results such as Theorems 12 and 15 suggest that pre-vote negotiations are a powerful tool players have to overcome the inefficiencies of aggregation rules. More specifically, when the goals of all players can be satisfied at the same time, pre-vote negotiations allow players to engineer side-payments—essentially as devices for credible commitments—leading to equilibrium outcomes that satisfy them, and ruling out all the others. We stress that in solving endogenous aggregation games, even when the game ends up sustaining efficient outcomes, players' strategies are individually rational and the game remains non-cooperative throughout. This differentiates our work from approaches to equilibrium selection with coalitional games, such as the one developed by Bachrach et al. (2011).

5. Lifting the Cube Assumption on Goals

An inspection of our main results of Section 4 should make it immediately clear that we have crucially relied on the assumption that agents' goals are propositional formulas known as cubes (Definition 1). Cubes guarantee a clear relation between truthfulness and weak dominance in aggregation games (recall Examples 1 and 2). While a direct generalisation of our main results is out of reach, in this section we show that a larger class of goals still support useful, albeit weaker, conclusions in the form of guaranteeing the existence of at least one surviving "good" equilibrium.

5.1 Monotonic Goals

As shown in Examples 1 and 2, when goals are not cubes it is possible to construct constant aggregation games in which truthful strategies are not dominant. In this section we find sufficient conditions on the class of goals for which there exists a truthful strategy that is weakly dominant, and we build further on this result by showing that, in constant aggregation games where players' goals are 'aligned' in a precise sense, we can still construct a truthful SNE that is also efficient (Theorem 18). Up until now we referred to aggregation games assuming players' goals were cubes (Definition 1). In this section we need to drop such constraint in the definition and, when referring to aggregation games, we make explicit what type of goal formulas we are considering.

We first need some additional notation and a definition. Let B be a valuation over PS, denote by B_{-j} the restriction of B to $PS \setminus \{p_j\}$.

Definition 10. *A formula $\varphi \in \mathcal{L}_{PS}$ is:*

positively monotonic *in p_j, where p_j occurs in φ, if for all $B, B' \in \mathcal{D}$ such that $B \models \varphi$, $B_{-j} = B'_{-j}$, and $B' \models p_j$ then $B' \models \varphi$.*

negatively monotonic *in p_j if for all $B, B' \in \mathcal{D}$ such that $B \models \varphi$, $B_{-j} = B'_{-j}$, and $B' \not\models p_j$ then $B' \models \varphi$.*

*A formula φ is **monotonic** if for every p_j occurring in φ, it is either positively or negatively monotonic in p_j.*

Let us mention a few examples. Formulas that are satisfied by only one model are monotonic. Similarly, cubes (i.e., conjunctions of literals) are monotonic. Not all monotonic formulas are, however, cubes: disjunctions of literals, which were used to construct Example 1, are monotonic but are not cubes in general. Again, Example 2 provides a good example of formulas which are not monotonic in the above sense. When viewed as characteristic functions, monotonic constraints correspond to the widely studied class of *unate boolean functions* (McNaughton, 1961).

5.2 Existence of Truthful and Efficient SNE

We have shown (Proposition 4) that when goals are cubes, every truthful strategy in a constant aggregation game is weakly dominant. Although this fails to be true when goals are monotonic (Example 1 offers a counterexample to such claim), we can still establish the following weaker result:

Lemma 16. *Let $\mathcal{A}$ be a constant aggregation game with monotonic goals, defined for a non-manipulable aggregator F. For each player $i \in \mathcal{N}$, if γ_i is consistent then there exists a truthful strategy for i that is weakly dominant in $\mathcal{A}$.*

Proof. The proof is by construction. Recall Definition 2. We build a ballot B_i such that $B_i \models \gamma_i$ and show that for any ballot $B'_i \neq B_i$ and profile $\boldsymbol{B}$, we have that if $F(\boldsymbol{B}_{-i}, B'_i) \models \gamma_i$ then $F(\boldsymbol{B}_{-i}, B_i) \models \gamma_i$. Let ballot B_i be as follows. For any variable p_j occurring in γ_i, let:

$$B_i(j) = \begin{cases} 1 & \text{if } \gamma_i \text{ is positively monotonic in } p_j \\ 0 & \text{if } \gamma_i \text{ is negatively monotonic in } p_j \\ 0 & \text{if } p_j \text{ does not occur in } \gamma \end{cases}$$

By the assumption of monotonicity of γ_i, together with its consistency, the above construction guarantees that $B_i \models \gamma_i$. To simplify the proof, since variables not occurring in γ_i do not play a role in its satisfaction, we can safely assume that all variables p_j for $j \in \mathcal{I}$ occur in γ_i. We now show that for any $B'_i \neq B_i$ we have that if $F(\boldsymbol{B}_{-i}, B'_i) \models \gamma_i$ then $F(\boldsymbol{B}_{-i}, B_i) \models \gamma_i$ as well, for any ballot profile $\boldsymbol{B}$, thereby establishing the claim. So take such a ballot B'_i and assume that $F(\boldsymbol{B}_{-i}, B'_i) \models \gamma_i$. Now let n be the Hamming distance[29] between B'_i and B_i and consider the sequence $B'_i = B^0_i, \ldots, B^n_i = B_i$ such that for all $1 \leq k \leq n$, B^{k+1}_i is equal to B^k_i except for the value of one of the issues (that is, the sequence consisting of a minimal number of value swaps to turn B'_i into B_i). By construction of B_i, each B^k_i is the result of swaps $0 \mapsto 1$ for issues in which γ_i is positively monotonic, and $1 \mapsto 0$ for issues in which γ_i is negatively monotonic, and possibly other swaps on variables not occurring in γ_i.

We now prove that $F(\boldsymbol{B}_{-i}, B^k_i) \models \gamma_i$ for all $0 \leq k \leq n$. We proceed by induction over k. For $k = 0$ the claim is true by assumption. Assume the claim is true for k (IH), i.e., $F(\boldsymbol{B}_{-i}, B^k_i) \models \gamma_i$. By the observation above, B^{k+1}_i can be obtained from B^k_i in only two

29. We recall that the hamming distance $H(B, B')$ between two valuations (vectors) B and B' is defined as follows: $H(B, B') = \sum_j |B(j) - B'(j)|$.

ways: [a] by a swap $0 \mapsto 1$ for some issue j in which γ_i is positively monotonic; or [b] by a swap $1 \mapsto 0$ for some issue j in which γ_i is negatively monotonic. If [a] is the case then in profile $(\boldsymbol{B}_{-i}, B_i^{k+1})$ there is one more voter, namely i, who supports p_j. We distiguish two cases:

[a1] If $F(\boldsymbol{B}_{-i}, B_i^k) \models p_j$, by independence and monotonicity of F, we can infer that also $F(\boldsymbol{B}_{-i}, B_i^{k+1}) \models p_j$ and therefore $F(\boldsymbol{B}_{-i}, B_i^{k+1}) = F(\boldsymbol{B}_{-i}, B_i^k) \models \gamma_i$ by IH.

[a2] If $F(\boldsymbol{B}_{-i}, B_i^k) \not\models p_j$, then either i is not pivotal on issue j, and therefore $F(\boldsymbol{B}_{-i}, B_i^{k+1}) = F(\boldsymbol{B}_{-i}, B_i^k) \models \gamma_i$ by IH. Or i is pivotal at step k, and then $F(\boldsymbol{B}_{-i}, B_i^{k+1}) \models p_j$. We now need to observe that by the assumption we have that γ_i is positively monotonic in p_j to conclude that, also in this case, $\mathsf{maj}(\boldsymbol{B}_{-i}, B_i^{k+1}) \models \gamma_i$.

This proves the claim for the first case. If [b] is the case, we reason in a symmetric fashion to conclude that that $F(\boldsymbol{B}_{-i}, B_i) \models \gamma_i$, as claimed. □

A direct consequence of Lemma 16 is the existence of a truthful NE in weakly dominant strategies for constant aggregation games with monotonic goals. However, in order to obtain results analogous to Theorem 12 in the context of monotonic goals we need first to establish sufficient conditions for the existence of truthful *and* efficient NE.

Let us first introduce some auxiliary terminology. Two monotonic goals γ_1 and γ_2 are **aligned** if they are positively monotonic on the same set of issues and negatively monotonic on the same set of issues. We establish the following result:

Lemma 17. *Let $\mathcal{A}$ be a constant aggregation game with monotonic goals, for a non-manipulable aggregator F. Assume moreover that $\mathcal{A}$ is C-consistent, that $C \in \mathcal{W}_{\bigwedge \Gamma}$ where $\Gamma = \{\gamma_i \mid i \in C\}$, and that all goals of agents $i \in C$ are aligned. Then, $\mathcal{A}$ has a truthful and C-efficient NE (in weakly dominant strategies).*

Proof. Let B'' be a ballot such that $B'' \models \bigwedge_{i \in C} \gamma_i$. B'' exists by assumption of C-consistency. B'' can be used in the proof of Lemma 16 to construct a truthful NE in weakly dominant strategies. Now observe that, since all goals of agents in C are aligned, the weakly dominant equilibrium so constructed is composed by a unanimous ballot choice for individuals in C, which we shall call B^*. Since C is a winning coalition for $\bigwedge_{i \in C} \gamma_i$, we have that $F(\boldsymbol{B}) \models \bigwedge_{i \in C} \gamma_i$ and the equilibrium is therefore also C-efficient as claimed. □

Everything is now in place to prove a variant of Theorem 12 for constant aggregation games where players hold monotonic goals.

Theorem 18. *Let $\mathcal{A}$ be a constant $\mathcal{N}$-consistent aggregation game with monotonic goals, for a non-manipulable aggregator F. Assume that all individual goals are aligned, and that $\mathcal{N} \in \mathcal{W}_{\bigwedge \Gamma}$ for $\Gamma = \{\gamma_i \mid i \in \mathcal{N}\}$. Then, there exists an $\mathcal{N}$-truthful and $\mathcal{N}$-efficient NE of $\mathcal{A}$ which is also a SNE of $\mathcal{A}$.*

Sketch of proof. Let $\boldsymbol{B}$ be an $\mathcal{N}$-truthful and $\mathcal{N}$-efficient NE of $\mathcal{A}$, which exists by Lemma 17. We can therefore construct a transfer profile τ such that $\boldsymbol{B}$ is a weakly dominant strategy equilibrium of $\tau(\mathcal{A}^{\mathcal{T}})$. The construction proceeds in the same way as the construction, for a transfer profile of the same type, given in the proof of Lemma 11. We then have to show

that $\boldsymbol{B}$ is a SNE of $\mathcal{A}^{\mathcal{T}}$. To do so we proceed towards a contradiction and suppose that there exists a profitable deviation τ_i^* for some player i in $\mathcal{A}^{\mathcal{T}}$. The argument used in the proof of Theorem 12 to the effect that no such profitable deviation exists can be applied directly, thereby establishing the claim. □

The assumption of monotonicity cannot be further weakened. If goals are allowed to be non-monotonic, then it is possible to construct (constant) aggregation games where, for some player, no truthful strategy is weakly dominant as witnessed by Example 1.

6. Discussion and Related Work

In this section we discuss our results from two points of view. First, we show how our results can be applied to the preservation of logical consistency when aggregation occurs on logically interconnected issues, which is the key problem of judgment aggregation. Second, we relate aggregation games to the influential notion of boolean game.

6.1 Pre-vote Negotiations and Collective Consistency

We showcase an application of endogenous aggregation games to binary aggregation with constraints, or judgment aggregation (Endriss, 2016; Grossi & Pigozzi, 2014), where individual ballots need to satisfy a logical formula, the *integrity constraint*, in order to be considered feasible or admissible. In case each individual provides an admissible ballot, the obvious question is whether the outcome of a given aggregation rule will be admissible, as well. Here is an instance of this problem.

Example 6. *Consider the scenario in Table 1. Suppose we impose the integrity constraint* $p \to (q \vee r)$*, making ballot* $(1, 0, 0)$ *inadmissible. All individual ballots in the example satisfy this requirement but the majority ballot does not.*

Paradoxical situations as those in Example 6 can be viewed as undesirable outcomes of aggregation games. Building on the example, assume that each agent has the following goals: $\gamma'_A = p, \gamma'_B = q, \gamma'_C = \neg r$. Let $\pi_A = \pi_B = \pi_C$ be constant payoff functions. Observe that parties' goals are all consistent with the integrity constraint $r \to (p \vee q)$, and that the admissible ballot $(1, 1, 0)$ satisfies all of them. Given these goals, the profile in Table 1 shows a truthful NE that, however, does not satisfy neither the goal of party B nor the integrity constraint $p \to (q \vee r)$. However, this equilibrium is *not* surviving because, intuitively, party B could transfer enough utility to party C for it to vote for q.

But the key question is whether we can guarantee that inadmissible equilibria do not survive.[30] It is easy to see that if the integrity constraint is implied by some player's goal—intuitively, the player internalises consistency itself as a goal—then $\mathcal{N}$-truthful and $\mathcal{N}$-efficient equilibria will satisfy the integrity constraint and, by our results, they will be surviving in games with well-behaved aggregators (Theorem 12). Vice versa, since only equilibria survive which are efficient for some winning coalition, with some extra assumptions on the aggregator (Theorem 15), only collectively consistent outcomes are generated by the aggregation.

30. That voting paradoxes can be studied from an equilibrium refinement perspective is an old but rather underexplored idea (Gueth & Selten, 1991).

6.2 Boolean Games

Boolean games (Harrenstein et al., 2001) are a logic-based representation of strategic interaction, where a set of individuals $A = \{1, 2, \ldots, m\}$ is assigned control of a set of propositional variables $P = \{p_1, p_2, \ldots, p_k\}$. Specifically, propositions are partitioned among the agents, i.e., each agent is assigned unique control over a subset of them, and each agent can decide to set the propositional variables he or she controls to true or false, in such a way that the final outcome of the boolean game is determined by the agents' truth value assignment on the variables each of them controls. Finally, each agent is equipped with a goal formula, i.e., a formula of propositional logic over the set of variables P. Typically, although agents have control over some propositional variables, they might not be able to realise their goal formula on their own.

Boolean games as aggregation games Boolean games can be seen as a special case of aggregation games, where each indiviudal is a dictator for the variables he or she controls. That is, a boolean game $\mathcal{B}$, defined over A and PS, can be seen as an aggregation game of the following form: $\mathcal{A}^{\mathcal{B}} = \left\langle \mathcal{N}^{\mathcal{B}}, \mathcal{I}^{\mathcal{B}}, F^{\mathcal{B}}, \{\gamma_i^{\mathcal{B}}\}_{i \in \mathcal{N}}, \{\pi_i^{\mathcal{B}}\}_{i \in \mathcal{N}} \right\rangle$ where: $\mathcal{N}^{\mathcal{B}} = A$ is the set of players, $\mathcal{I}^{\mathcal{B}} = PS$ is the set of issues, $F^{\mathcal{B}}$ is a dictatorship on issue j by i, for any issue $j \in \mathcal{I}$ controlled by $i \in N$, i.e., $F^{\mathcal{B}}(\boldsymbol{B})(j) = B_i(j)$, $\gamma_i^{\mathcal{B}}$ is the goal formula for i in $\mathcal{B}$, and each $\pi_i^{\mathcal{B}}$ is constant.

Thanks to the above reduction we are able to import the following complexity bounds from the boolean games literature:

Proposition 19. *Let $\mathcal{A}$ be an aggregation game with arbitrary goals γ_i for each $i \in \mathcal{N}$ and $\boldsymbol{B}$ a ballot profile.*

(1) The problem of verifying whether, for some transfer profile τ, $\boldsymbol{B}$ is a NE of $\tau(\mathcal{A})$ is co-NP-hard;

(2) The problem of verifying whether there exists a transfer function τ and a ballot profile $\boldsymbol{B}'$ that is a Nash equilibrium of $\tau(\mathcal{A})$ and such that $F(\boldsymbol{B}') \models \bigwedge_{i \in N} \gamma_i$ is Σ_p^2-hard;

(3) The problem of verifying whether there exists a transfer function τ such that for all ballot profiles $\boldsymbol{B}'$ that are a Nash equilibrium of $\tau(\mathcal{A})$ we have that $F(\boldsymbol{B}') \models \bigwedge_{i \in N} \gamma_i$ is Σ_p^2-hard.

Proof. The results follow from (Wooldridge et al., 2013, Proposition 1), (Wooldridge et al., 2013, Proposition 6) and (Wooldridge et al., 2013, Proposition 14) respectively, together with the translation given above. □

Boolean games have been extended in a number of ways, some of which will be dealt with next. It is however worth mentioning the extension by Gerbrandy (2006) and Belardinelli, Grandi, Herzig, Longin, Lorini, Novaro, and Perrussel (2017) to structures where coalitions are able to share the control of a propositional variable. Although the purpose is to study the logical properties of shared control and not the property of social choice functions, Gerbrandy (2006) basically works with aggregation games with arbitrary aggregators, but without goals and without payoff function. The work of Belardinelli et al. (2017) takes instead a verification approach, showing that games with shared control can be simulated by

classical boolean games for what concern the model-checking of alternating-time temporal logic formulas. They do not, however, study the game-theoretic structure of these models.

Boolean games and incentive engineering A class of boolean games that is relevant to our framework is boolean games with arbitrary payoffs, i.e., aggregation games of the form $\mathcal{A}^{\mathcal{B}} = \left\langle \mathcal{N}^{\mathcal{B}}, \mathcal{I}^{\mathcal{B}}, F^{\mathcal{B}}, \{\gamma_i^{\mathcal{B}}\}_{i\in\mathcal{N}}, \{\pi_i^{\mathcal{B}}\}_{i\in\mathcal{N}} \right\rangle$ where the payoff is not necessarily constant. These boolean games have been introduced to account for efforts (or costs) in performing actions (Grant, Kraus, Wooldridge, & Zuckerman, 2011; Wooldridge et al., 2013; Turrini, 2016; Harrenstein et al., 2014). When comparing two outcomes, a player will prefer the ones satisfying the goal, but will otherwise look at minimising the effort. This amounts to the same idea of having a payoff that is taken into account only in case goal satisfaction cannot discriminate between outcomes.

Boolean games with costs have been looked at from the point of view of *incentive engineering*, allowing payoffs to be manipulable, either by exogenous taxation mechanisms as in the work of Wooldridge et al. (2013) and Harrenstein et al. (2014), or by endogenous negotiation as in the work of Turrini (2016). In the exogenous setting, an external system designer can impose taxes on players' actions, by effectively influencing their decision-making towards the realisation of her own goal formula. In the endogenous setting, individuals undergo a pre-play negotiation phase and try to improve upon their final payoff using side-payments. This has lead to the discovery of the existence of *hard equilibria* (Harrenstein et al., 2014), i.e., pure Nash-equilibria that cannot be removed by external incentives. Note that their presence is even more frequent in endogenous settings, due to the fact that side-payments are a weaker form of manipulation than external intervention, as observed by Turrini (2016). Endogenous boolean games are essentially endogenous aggregation games applied to a restricted setting. However the idea of hard equilibrium does carry over to aggregation games in general. Think for instance of a situation in which there is only one issue, p, and only one winning coalition, $\mathcal{N}$. If everyone wants p to be true, then the profile in which everyone votes for p is an equilibrium that is impossible to remove by manipulating payoffs.

7. Conclusions

The paper has proposed a model of pre-vote negotiation for games of binary aggregation. Although a number of papers in the literature on voting games have focused on the problem of avoiding undesirable equilibria, no model studying strategic behaviour in a pre-vote negotiation phase has so far been proposed. We used our model to show how pre-vote negotiations can restore the efficiency of truthful voting in such games. More specifically we established the following main results. In uniform aggregation games for non-manipulable (i.e., independent and monotonic) aggregators where voters' goals are cubes, if an equilibrium is truthful and efficient it will be selected by equilibrium behaviour in the pre-vote negotiation phase (Theorem 12). We also showed that the 'only if' variant of this claim holds with respect to the efficiency alone of the equilibria (Theorem 15). For larger classes of goals we were able to show the existence of a truthful and efficient equilibrium that will be selected by equilibrium behaviour in the pre-vote negotiation phase (Theorem 18, for monotonic goals) and that this assumption cannot be further relaxed (Example 1).

Our work is a first step towards the development of a body of theoretical results on how strategic interaction preceding voting influences the outcomes of group decision-making. While classical social choice theory analyses individual preferences as independent inputs for an aggregation problem, we have shown how the introduction of an explicit negotiation phase preceding the vote has a fundamental impact in equilibrium selection. In particular, we have seen how the mere possibility of pre-vote negotiation rules out the selection of unintuitive and undesirable equilibria in which voters have aligned motives but choose to go against them together. A number of directions are possible to further enrich our framework, with two main ones. First, the analysis of further co-dependence among individuals, by for instance studying interactions constrained in a social network (Jackson, 2008), where individuals can actively incentivise their connections only. Second, the elaboration of an even more realistic model of pre-vote negotiation, studying resource-bounded individuals that have an upper bound, in the form of an endowment, on the amounts they can transfer in order to influence each other's behaviour.

Acknowledgments

This paper improves and extends work previously presented at the 24th International Joint Conference on Artificial Intelligence, IJCAI'15 (Grandi, Grossi, & Turrini, 2015). The paper has benefited from the feedback of the anonymous reviewers at IJCAI'15, COMSOC'14 and LOFT'14. The authors are greatly indebted to Edith Elkind, Ulle Endriss and the anonymous reviewers of JAIR for valuable comments on earlier versions of this work. The authors also wish to thank the participants of the 2nd International Workshop on Norms, Actions and Games (NAG'2016) in Toulouse, the 2015 LABEX CIMI Pluridisciplinary Workshop on Game Theory in Toulouse, the 5th International Workshop on Computational Social Choice (COMSOC'2014) in Pittsburgh, the 11th Conference on Logic and the Foundations of Game and Decision Theory (LOFT'2014) in Bergen, the 12th Meeting of the Social Choice and Welfare Society held in Boston in 2014, and the Workshop on Fair Division, Voting and Computational Complexity held in Graz in 2014, where this work has been presented. Davide Grossi acknowledges partial support for this research by EPSRC under grant EP/M015815/1. Paolo Turrini acknowledges support from Imperial College London under the Imperial College Research Fellowship "Designing negotiation spaces for collective decision-making" (DoC AI1048).

References

Bachrach, Y., Elkind, E., & Faliszewski, P. (2011). Coalitional voting manipulation. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*.

Baumeister, D., Erdélyi, G., Erdélyi, O. J., & Rothe, J. (2015). Complexity of manipulation and bribery in judgment aggregation for uniform premise-based quota rules. *Mathematical Social Sciences*, *76*, 19–30.

Belardinelli, F., Grandi, U., Herzig, A., Longin, D., Lorini, E., Novaro, A., & Perrussel, L. (2017). Relaxing exclusive control in boolean games. In *Proceedings of the 16th conference on Theoretical Aspects of Rationality and Knowledge (TARK)*.

Benôit, J.-P., & Kornhauser, L. (2010). Only a dictatorship is efficient. *Games and Economic Behavior*, *70*(2), 261–270.

Botan, S., Novaro, A., & Endriss, U. (2016). Group manipulation in judgment aggregation. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Brams, S., & Fishburn, P. (1978). Approval voting. *American Political Science Review*, *72*(3), 831–847.

Brams, S. J., Kilgour, D. M., & Zwicker, W. S. (1998). The paradox of multiple elections. *Social Choice and Welfare*, *15*(2), 211–236.

Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. (Eds.). (2016). *Handbook of Computational Social Choice*. Cambridge University Press.

Chopra, S., Ghose, A., & Meyer, T. (2006). Social choice theory, belief merging and strategy-proofness. *Information Fusion*, *7*(1), 61–79.

Desmedt, Y., & Elkind, E. (2010). Equilibria of plurality voting with abstentions.. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC)*.

Dietrich, F., & List, C. (2007a). Arrow's theorem in judgment aggregation. *Social Choice and Welfare*, *29*(19–33).

Dietrich, F., & List, C. (2007b). Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics*, *19*(4), 391–424.

Dietrich, F., & List, C. (2007c). Strategy-proof judgment aggregation. *Economics and Philosophy*, *23*(3), 269–300.

Dokow, E., & Holzman, R. (2010). Aggregation of binary evaluations. *Journal of Economic Theory*, *145*(2), 495–511.

Dryzek, J., & List, C. (2003). Social choice theory and deliberative democracy: A reconciliation. *British Journal of Political Science*, *33*, 1–28.

Elkind, E., Grandi, U., Rossi, F., & Slinko, A. (2015). Gibbard-satterthwaite games. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*.

Elkind, E., & Lackner, M. (2015). Structure in dichotomous preferences. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*.

Endriss, U. (2016). Judgment aggregation. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (Eds.), *Handbook of Computational Social Choice*, chap. 17. Cambridge University Press.

Endriss, U., Grandi, U., & Porello, D. (2012). Complexity of judgment aggregation. *Journal of Artificial Intelligence Research*, *45*, 481–514.

Everaere, P., Konjeczny, S., & Marquis, P. (2007). The strategy-proofness landscape of merging. *Journal of Artificial Intelligence Research*, *28*(1), 49–105.

Faliszewski, P., & Rothe, J. (2016). Control and bribery in voting. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. (Eds.), *Handbook of Computational Social Choice*, chap. 7, pp. 146–168. Cambridge University Press.

Gerbrandy, J. (2006). Logics of propositional control. In *Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Goranko, V., & Turrini, P. (2016). Two-player preplay negotiation games with conditional offers. *International Game Theory Review*, *18*(1), 1–31.

Grandi, U., & Endriss, U. (2013). Lifting integrity constraints in binary aggregation. *Artificial Intelligence*, *199-200*, 45–66.

Grandi, U., Grossi, D., & Turrini, P. (2015). Equilibrium refinement through negotiation in binary voting. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*.

Grant, J., Kraus, S., Wooldridge, M., & Zuckerman, I. (2011). Manipulating boolean games through communication. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*.

Grossi, D., & Pigozzi, G. (2014). *Judgment Aggregation: A Primer*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers.

Gueth, W., & Selten, R. (1991). Majority voting in the Condorcet paradox as a problem of equilibrium selection. In *Game Equilibrium Models IV*. Springer Berlin Heidelberg.

Halpern, J. (2011). Beyond Nash-equilibrium: Solution concepts for the 21st century. In Apt, K., & Grädel, E. (Eds.), *Lectures in Game Theory for Computer Scientists*, pp. 264–289. Cambridge University Press.

Harrenstein, P., Turrini, P., & Wooldridge, M. (2014). Hard and soft equilibria in boolean games. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

Harrenstein, P., Turrini, P., & Wooldridge, M. (2017). Characterising the manipulability of boolean games. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*.

Harrenstein, P., van der Hoek, W., Meyer, J., & Witteveen, C. (2001). Boolean games. In *Proceeding of the 8th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*.

Hazon, N., Lin, R., & Kraus, S. (2013). How to change a group's collective decision?. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)*.

Hodge, J. (2002). *Separable Preference Orders*. Ph.D. thesis, Western Michigan University.

Jackson, M. O. (2008). *Social and Economic Networks.* Princeton University Press.

Jackson, M. O., & Wilkie, S. (2005). Endogenous games and mechanisms: Side payments among players. *Review of Economic Studies*, *72(2)*, 543–566.

Kim, Y. (1996). Equilibrium selection in n-person coordination games. *Games and Economic Behavior*, *15*, 203–227.

Konieczny, S., & Pino Peréz, R. (2002). Merging information under constraints: A logical framework. *Journal of Logic and Computation*, *12*(15), 773–808.

Lacy, D., & Niou, E. M. S. (2000). A problem with referendums. *Journal of Theoretical Politics*, *12*(1), 5–31.

Lang, J., & Xia, L. (2016). Voting over combinatorial domains. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. (Eds.), *Handbook of Computational Social Choice*. Cambridge University Press.

List, C. (2011). Group communication and the transformation of judgments: An impossibility result. *The Journal of Political Philosophy*, *19*(1), 1–27.

McNaughton, R. (1961). Unate truth functions. *IRE Transactions on Electronic Computers*, *EC-10*(1), 1–6.

Meir, R. (2017). Iterative voting. In Endriss, U. (Ed.), *Trends in Computational Social Choice*, chap. 4. AI Access.

Meir, R., Lev, O., & Rosenschein, J. S. (2014). A local-dominance theory of voting equilibria. In *Proceedings of the 15th ACM Conference on Economics and Computation (EC)*.

Meyerson, R. B. (1978). Refinements of the Nash equilibrium concept. *International Journal of Game Theory*, *7*(2), 73–80.

Monderer, D., & Tennenholtz, M. (2004). K-implementation. *Journal of Artificial Intelligence Research*, *21*(1), 37–62.

Monderer, D., & Tennenholtz, M. (2009). Strong mediated equilibrium. *Artificial Intelligence*, *173*(1), 180 – 195.

Obraztsova, S., Markakis, E., & Thompson, D. R. M. (2013). Plurality voting with truth-biased agents.. In *Proceedings of the 6th International Symposium on Algorithmic Game Theory (SAGT)*.

Obraztsova, S., Rabinovich, Z., Elkind, E., Polukarov, M., & Jennings, N. R. (2016). Trembling hand equilibria of plurality voting. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*.

Ozkai-Sanver, I., & Sanver, M. (2006). Ensuring Pareto optimality by referendum voting. *Social Choice and Welfare*, *27*(1), 211–219.

Rubinstein, A. (2012). *Lecture Notes in Microeconomic Theory: The Economic Agent*. Princeton University Press.

Shoham, Y., & Leyton-Brown, K. (2008). *Multiagent Systems: Algorithmic, Game-Theoretic and Logical Foundations*. Cambridge University Press.

Turrini, P. (2016). Endogenous games with goals: side-payments among goal-directed agents. *Autonomous Agents and Multi-Agent Systems*, *30*(5), 765–792.

Wilson, R. (1975). On the theory of aggregation. *Journal of Economic Theory*, *10*(1), 89–99.

Wooldridge, M., Endriss, U., Kraus, S., & Lang, J. (2013). Incentive engineering in boolean games. *Artificial Intelligence*, *195*, 418–439.

Xia, L., & Conitzer, V. (2010). Stackelberg voting games: Computational aspects and paradoxes. In *Proceedings of the 24th conference on Artificial Intelligence (AAAI)*.